Chapter 1

Introduction
Chapter 1. Introduction

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1 GENERAL MATTER</td>
<td>3</td>
</tr>
<tr>
<td>1.1.1 Thesis Objectives</td>
<td>7</td>
</tr>
<tr>
<td>1.1.2 Thesis Overview</td>
<td>8</td>
</tr>
<tr>
<td>1.2 PHARMACEUTICAL PRODUCT DEVELOPMENT</td>
<td>10</td>
</tr>
<tr>
<td>1.2.1 An Overview</td>
<td>10</td>
</tr>
<tr>
<td>1.2.2 In Vitro Screening and Animal Models</td>
<td>13</td>
</tr>
<tr>
<td>1.2.3 In Silico Screening</td>
<td>16</td>
</tr>
<tr>
<td>1.3 MODELING TECHNIQUES IN PHARMACOKINETICS</td>
<td>19</td>
</tr>
<tr>
<td>1.3.1 Non Structure-Based methods</td>
<td>20</td>
</tr>
<tr>
<td>1.3.2 Structure-Based Methods</td>
<td>22</td>
</tr>
<tr>
<td>1.3.2.1 Multilinear Regression</td>
<td>23</td>
</tr>
<tr>
<td>1.3.2.2 Artificial Neural Networks</td>
<td>26</td>
</tr>
<tr>
<td>1.4 ARTIFICIAL INTELLIGENCE SYSTEMS</td>
<td>28</td>
</tr>
<tr>
<td>1.4.1 Multilayer Perceptron ANNs</td>
<td>29</td>
</tr>
<tr>
<td>1.4.2 Radial-Basis Function ANNs</td>
<td>32</td>
</tr>
<tr>
<td>1.4.3 Other Soft Computing Methods</td>
<td>34</td>
</tr>
<tr>
<td>1.4.4 Descriptor Selection</td>
<td>36</td>
</tr>
<tr>
<td>1.5 DESCRIPTORS USED IN MODELING</td>
<td>39</td>
</tr>
<tr>
<td>1.5.1 Constitutional Descriptors</td>
<td>40</td>
</tr>
<tr>
<td>1.5.2 Topological Indices</td>
<td>42</td>
</tr>
<tr>
<td>1.5.2.1 Connectivity Indices</td>
<td>43</td>
</tr>
<tr>
<td>1.5.2.2 Electrotopological Indices</td>
<td>45</td>
</tr>
<tr>
<td>1.5.3 Quantum Chemical Numbers</td>
<td>46</td>
</tr>
<tr>
<td>1.5.4 Solubility and Partitioning</td>
<td>48</td>
</tr>
<tr>
<td>1.5.5 Other Descriptors</td>
<td>49</td>
</tr>
<tr>
<td>1.6 STRUCTURE-PHARMACOKINETIC RELATIONSHIPS</td>
<td>50</td>
</tr>
<tr>
<td>1.6.1 Absorption</td>
<td>51</td>
</tr>
<tr>
<td>1.6.2 Distribution</td>
<td>56</td>
</tr>
<tr>
<td>1.6.3 Metabolism and Excretion</td>
<td>60</td>
</tr>
<tr>
<td>1.6.3.1 Clearance</td>
<td>63</td>
</tr>
<tr>
<td>1.7 SUMMARY REMARKS</td>
<td>65</td>
</tr>
</tbody>
</table>
Chapter 1. Introduction

1.1 General Matter

The technology and research revolution has provided many areas of science and industry with tools for more extensive and efficient operation. Nowhere is this phenomenon more evident than for new drug discovery and development in the pharmaceutical industry. Exploring the relationship between the structure of a molecule and its various biological and biochemical properties is the basis of drug discovery. Modern approaches to this field of study employ a combination of techniques. These include tests based on combinatorial chemistry and high-throughput (HT) screening as well as rational pharmaceutical design based on geometric and chemical characteristics of molecule-molecule interactions. Furthermore, understanding and optimising factors such as the effect of a compound on the body and the effect of the body on a compound are essential in developing a new drug.

The main bottleneck in drug discovery is the identification of new chemical entities (NCEs) to be used for drug leads. The 1990s saw development of new automated tools for drug discovery including combinatorial chemistry and high-throughput screening. These tools have led to the increased discovery of new drug lead compounds each of which in turn require pharmacological and pharmacokinetic testing. Moreover, substantial increases in computing power as well as development of robust software has given scientists the opportunity to
undertake significant research projects from their own desktops. Consequently, data analysis, data mining, and information manipulation have all benefited and progressed considerably.

Software programs have been developed for a wide range of fields such as quantitative structure-activity relationship (QSAR) analyses, pharmacophore elucidation, molecular modeling, drug-receptor interactions and *in vivo* simulations. Newer techniques have been influenced by what is termed “soft computing” which aims to accommodate the imprecision and uncertainty inherent in the real world [Zadeh, 1996]. Soft computing draws on the model of the human brain and derives mainly from artificial intelligence (AI) sources including genetic algorithm (GA), fuzzy logic, and artificial neural network (ANN) approaches [Maddalena, 1998]. Other less common techniques include cellular automata, fractals and chaos theory. ANNs are particularly useful for modeling nonlinear systems. Although not as common in the pharmaceutical industry as conventional modeling and mathematical techniques, soft computing has been successful in a number of fields in the industry.

In particular, soft computing has been useful in the development of quantitative structure-activity relationship and quantitative structure-property relationship (QSPR) models. General methods involve correlation of physicochemical descriptors of chemical compounds with either an activity or property value. Classically, one or two descriptors such as octanol/water partition coefficients and molar refractivity are experimentally determined for a group of congenic compounds [Hansch et al., 1995]. These experimental descriptor values are then related to some sort of biological activity. The result is a mathematical model
which describes the contribution of the descriptors to the activity. Once a predictive model has been built, numerous new potential-drug molecules which are chemically similar to those of the benchmark data set can then be screened from large databases. These molecules are all evaluated for their biological properties based on the predictive model developed. The aim is to target a few novel molecules with potentially attractive pharmaceutical properties that can then be tested further in the traditional way in the laboratory. Effective data mining techniques are vital to extract the information necessary to select these novel molecules.

Such models have used both whole molecule descriptors and descriptors for individual substitution positions and functional groups on structurally related compounds. Thus, for a single study there may be a large number of potential descriptors for correlation with only a single target activity or property. Not all descriptors are useful so selection of meaningful descriptors is crucial for successful model development.

Recently, theoretical descriptors generated only from the molecular structure of a compound have become popular. Over a thousand of these descriptors have been defined to date although not all are entirely useful [Balaban & Ivanciuc, 1999]. Many of these descriptors have been successfully correlated with parameters such as boiling points of alkanes [Cherqaoui & Villemín, 1994], aqueous solubility [Huuskonen et al., 1997], binding affinities [Beck et al., 1996], and analgesic properties [Galvez et al., 1994b].
Chapter 1. Introduction

Similarly, some models have also been constructed for drugs and their pharmacokinetic parameters [Seydel & Schaper, 1981; Hinderling et al., 1984a; Fouchecourt et al., 2001]. Termed quantitative structure-pharmacokinetic relationship (QSPkR) models, they are not as common in the literature as other QSPRs are. This has largely been attributed to the complex factors involved in some pharmacokinetic parameters such as hepatic metabolism and elimination half life, as well as the time dependency of drug concentration in vivo [Mayer & van de Waterbeemd, 1985].

Development of predictive QSPkR models would be of great interest to the pharmaceutical industry since it would allow valuable information to be gained very early during the drug development process. Should successful models be based on structure alone then predictions could be made for theoretical chemical structures before they need even be synthesised. Additionally, knowledge of human pharmacokinetics prior to clinical trials would enable decisions to be made regarding viability of potential drugs for continued development. This would impart another substantial time- and cost-saving benefit.

This introduction lays a foundation for the work presented in this thesis by presenting a survey of the relevant published literature. First, an overview of the drug development process is given with emphasis on where the present technique will be most beneficial. Current methods of screening are then described including in vitro, in vivo, and in silico techniques.

Next, various pharmacokinetic modeling approaches are explored so that comparisons between them can be made. Here, structure-based methods are
introduced as the method of choice in the present work. The different soft computing techniques used in structure-based modeling are presented next, with particular focus on the ANN paradigms used in the present work. Then, since structure-based pharmacokinetic modeling requires information in the form of descriptors, the importance of descriptor selection is briefly outlined.

Given that this modeling technique is data-driven, the nature and meaning of descriptors selected in models is very important. Hence, an explanation of relevant descriptors is provided after descriptor selection is described. Following this, the chapter presents a review of published QSPkR studies arranged according to pharmacokinetic properties. The pharmacokinetic parameters are dealt with as modeling objectives and cited literature are explored as they pertain to the present study. Modeling techniques, input variables used, data sets, and study outcomes are compared and contrasted to provide a background for the work presented in subsequent chapters.

1.1.1 Thesis Objectives

The broad aim of this thesis was to develop predictive QSPkR models using ANNs. Specific objectives were to:

- Identify various theoretical descriptors generated from molecular structure and examine their relevance to QSPkR studies.

- Investigate both selection of descriptors and effects of ANN architecture on QSPkR model performance.
• Explore relationships of theoretical descriptors with different pharmacokinetic parameters.

• Investigate the viability of multiple, simultaneous pharmacokinetic parameter prediction.

• Develop predictive QSPkR models for both structurally related and structurally unrelated sets of drugs.

1.1.2 Thesis Overview

Chapter 1 introduces the work and concepts presented in this thesis. The scope of the present research is rationalised by a review of the literature relevant to drug development and QSPkR analyses. Application of ANNs in pharmacy and the use of theoretical descriptors in QSAR and QSPR studies are also reviewed.

Chapter 2 describes the methods employed to obtain suitable chemical structures and the subsequent derivation of theoretical descriptors from them. Implementation of multilayer perceptron and radial-basis function ANNs is described along with the technical aspects behind these ANN paradigms. General procedures for QSPkR model construction are also outlined.

Chapter 3 carefully investigates the method of selective descriptor pruning in ANN studies. Simulated and real pruning exercises are performed to examine systematic removal of descriptors and the effect it has on optimum descriptor selection. The ratio of the number of patterns to the number of connections, \( \rho \), is also analysed to determine its importance in QSAR and QSPR studies.
Chapter 4 presents a QSPkR study based on the selective descriptor pruning technique described in Chapter 3. A multilayer perceptron ANN is used to develop models for a small set of $\beta$-adrenoceptor antagonists. Descriptors include connectivity and charge indices, topological indices, calculated log $P$ values, and simple constitutional descriptors. Separate models are constructed for each of a number of different pharmacokinetic parameters.

Chapter 5 presents a QSPkR study where multiple pharmacokinetic parameters for a series of cephalosporins are predicted simultaneously. Sensitivity-based pruning is used to select an optimum set of descriptors from amongst constitutional, topological, chemical, geometrical, and quantum chemical descriptors. Mathematical relationships between different pharmacokinetic parameters are examined with respect to the descriptors included in the final model.

Chapter 6 presents a predictive QSPkR model for bioavailability of a large number of structurally diverse compounds. Descriptors generated include constitutional, topological, chemical, geometrical, quantum chemical numbers, calculated bulk properties, and solubility parameters. The model is constructed using a radial-basis function ANN. A combination of manual and sensitivity-based pruning is utilised to select the optimum descriptor set. The contribution of optimum descriptors to drug bioavailability is analysed.

Chapter 7 reviews the findings of the preceding chapters and discusses their significance as a whole. Further discussion regarding the current and potential future directions of ANN use in the pharmaceutical industry is presented. Overall conclusions are made based on the results of the present research.
1.2 Pharmaceutical Product Development

Development of successful pharmaceutical products drives profit, which in turn permits further drug development. Profit from sales occurs only after a marketable product has been produced and its associated developmental costs have been surpassed. In addition, sales for a successful product must also account for the cost of unsuccessful compounds which have failed at some stage of the developmental process.

1.2.1 An Overview

The drug development process involves elements from both regulatory bodies and industry (Figure 1-1). The entire process spans pre-clinical laboratory research and development, through to clinical evaluation, and finally to post-marketing surveillance. It is in pre-clinical research, commonly referred to as Phase 0, that new drug entities are screened and developed for eventual clinical application to humans. Owing to rapid advances in areas such as computing technology, combinatorial chemistry, molecular and cell biology, and high-throughput screening techniques, strong progress has been made in identifying potential lead compounds. Although such techniques have provided an increased number of potential new drug entities, this has not necessarily translated to an increased number of drugs successfully reaching the marketplace [Grass & Sinko, 2001].

Of all the NCEs screened in Phase 0, only a small number ever progress beyond animal studies. From this small percentage it has been estimated that less than one quarter possess all the necessary pharmacokinetic and pharmacodynamic characteristics to successfully become marketable products. Increasing the number
Chapter 1. Introduction

of NCEs progressing to the clinical trial phases then substantially increases the number of failures at this late stage. The bulk of drug development spending can be attributed to these failures, and the total amount has been estimated to be around 75% of all monies spent on drug development. Hence, the focus of drug development has expanded more and more to include procedures aimed at identifying potential failures as well as successes [Ekins et al., 2000].

**Figure 1-1.** General sequence of events involved in drug development (adapted from [Lesko et al., 2000]). Black areas (■) represent time taken for regulatory processes, whilst clear areas represent scientific process normally associated with industry.

In the period 1968-1988, it was found that the major reasons for failure of NCEs in humans were unacceptable pharmacokinetics (~40% of failures) and lack of efficacy (~30% of failures) [Prentis et al., 1988]. Unacceptable pharmacokinetics can include poor absorption, distribution, metabolism, or excretion (ADME) characteristics. Poor pharmacokinetics can also manifest as lack of clinical
Chapter 1. Introduction

efficacy. Hence, the human pharmacokinetics of a compound play a key role in determining the suitability of an NCE for further development.

Screening for ADME properties and toxicity is usually performed both in vitro and with various animal models which are time-consuming and expensive [Norris et al., 2000]. Even then, results may not always accurately reflect the pharmacokinetics of a compound once it is administered to humans. Results gained from ADME screening are used to determine whether development of an NCE should continue to Phase 1 or not.

Early clinical testing in humans conducted in healthy subjects is aimed at determining tolerated dose size, initial pharmacokinetic profiles, candidate delivery systems, and the relationship between plasma concentrations and pharmacological effects [Peck et al., 1992]. Preclinical screening may indicate an NCE with suitable pharmacokinetic attributes, however, the majority of candidates do not succeed through the clinical testing phases.

Owing to financial pressures and the need for more accurate predictive methods, research into methods other than in vitro screening and animal models is a growing area. Computational techniques, often termed in silico methods, have begun to play a larger role in the drug discovery process. Previously impossible computational tasks have become a matter of routine with the ever-increasing power and availability of computers and software. In silico methods have gained popularity in virtual compound library screening [Walters et al., 1998], three-dimensional (3D) pharmacophore elucidation [Terfloth & Gasteiger, 2001], and QSPR analyses [Ekins et al., 2000]. They are cheaper and quicker than performing
in vitro and animal experiments, although they are not yet as acceptable to regulatory authorities. Data taken from previously published work or from objective sources such as molecular structure have allowed prediction of important pharmacokinetic properties such as intestinal absorption [Wessel et al., 1998], distribution parameters [Herman & Veng-Pedersen, 1994], and binding characteristics [Wagener et al., 1995; Loukas, 2001]. Both experimental and in silico techniques have advantages and disadvantages, and their roles in drug development will now be discussed briefly.

1.2.2 In Vitro Screening and Animal Models

The use of in vitro and animal models allows research to be performed on a much cheaper and faster scale than in humans. Furthermore, safety issues and ethical requirements for humans are also avoided. These methods aim to provide measurements regarding the potential NCEs which can then hopefully be correlated with human activity or pharmacokinetics.

Most drugs are developed for oral administration so bioavailability of a compound is an important factor to consider. Oral bioavailability is dependent sequentially upon dissolution in the gastrointestinal (GI) tract, absorption across the physical barrier of the GI membrane, and a first pass through the liver and lungs [Sietsema, 1989]. Absorption occurs via either paracellular or transcellular pathways. Hence, absorption is influenced by permeability, molecular size, and hydrogen bonding characteristics [Smith & van de Waterbeemd, 1999]. Although not a measure of bioavailability, absorption is the first step in delivering an oral dose of a drug.
Chapter 1. Introduction

One common technique for *in vitro* modeling of absorption is the Caco-2 monolayer system which is an immortalised human colon adenocarcinoma cell line. Caco-2 cells are enterocyte-derived cells possessing a microvillus surface and allow moderate to high-throughput screening of compounds. Another technique, Madin-Darby canine kidney (MDCK) cells in monolayer can be used in a similar functional manner to Caco-2 cells but do not require the 2-3 week culturing times Caco-2 cells do [Pelkonen et al., 2001].

Although both techniques allow reasonably fast screening of compounds, the major disadvantage is that they are considerably different from the situation *in vivo*. Compounding this are the inter-experiment and inter-laboratory variations seen with these *in vitro* techniques. Thus, cellular models can provide useful information but are not complete in themselves.

High-throughput screening is the rapid analysis of chemical libraries for biological activity. Compounds are screened using automated miniaturised assays which enable vast numbers of compounds to be tested in a short period of time [Inglese, 2002].

Combinatorial libraries contain from hundreds to millions of compounds for testing, so examining the entire chemical space available to most companies would not be affordable. Efforts have instead been aimed at reducing the chemical space to those compounds for which manufacture and further development is a feasible proposition [Gobbi & Poppinger, 1998]. Searching such large numbers of compounds has increased the number of drug-like “hits” exhibiting potential
biological activity. This has in turn placed more pressure on the subsequent step of assessing the potential NCEs for suitable pharmacokinetic properties.

_In vivo_ metabolism of test compounds is a serious problem in new drug development. Metabolism of compounds by various enzymes can also be screened using _in vitro_ high-throughput methods. Typically, hepatic microsomal or other liver or tissue homogenate preparations are incubated in 96-well plates with individual compounds in each well. Reaction times and conditions are completely controlled, and extent of metabolism is compared with a standard.

The aim to mimic _in vivo_ metabolism is not always fully achieved since drug metabolism is a multifactorial process which usually involves multiple pathways [Gaviraghi et al., 2001]. In addition, protein binding can limit hepatic extraction _in vivo_ and may not be accounted for using simply the _in vitro_ screen. The major drawback with microsomal preparations, however, is the inconsistency between preparations which can lead to variable results [Spalding et al., 2000].

Allometric scaling is another method of screening potential NCEs for human application. It is based on the premise that human and animal anatomical, physiological and biochemical characteristics are comparable [Feng et al., 2000]. Once pharmacokinetic parameters have been determined in animals they can then be mathematically related to human pharmacokinetic parameters. Individually, there is no single animal which can be relied upon to accurately predict human pharmacokinetics. Consequently, data must often come from several different species to construct a one predictive model [Hussain et al., 1993]. Correction terms can also be applied to allometric calculations to increase the accuracy of
Chapter 1. Introduction

models. Allometric scaling has been applied to structurally diverse compounds [Feng et al., 2000] and for various pharmacokinetic parameters [Jezequel, 1994]. Although allometric studies avoid experiments on humans, they are still subject to ethics approval and are expensive to run. As a result of the inherent differences between animals and humans, allometric models do not always lead to accurate predictions for a human clinical situation [Grass & Sinko, 2001].

1.2.3 In Silico Screening

Computational methods for ADME began with the classical QSAR models developed last century using lipophilicity [Hansch & Lien, 1968]. Even though data sets were small and comprised of structurally similar compounds, the hypothesis that metabolism and activity could be modeled in a quantitative fashion based on structural considerations was pioneering. From these beginnings, much progress has been made using purely computational methods for compound screening. The scope of QSAR has evolved to account for the spatial arrangements of atoms in a candidate molecule. Interaction of a molecule with a receptor or enzyme is dependant upon the 3D configuration and conformation of that molecule: good steric arrangement of atoms and functional groups allows a more positive interaction. Once a 3D pharmacophore has been generated for a known agonist or substrate, screening of potential candidates for that particular receptor or enzyme may then commence. Other computational techniques have been combined with 3D QSAR, for example comparative molecular field analysis (CoMFA) which estimates steric and electrostatic interactions between a molecule and target binding site, and the VolSurf/GRID procedure which calculates energy-
favourable sites around a molecule and converts them into selected molecular descriptors [Ekins et al., 2000].

One of the critical requirements for these in silico screening techniques is the availability of virtual libraries of compounds containing structures able to be screened. If the number of compounds available for in vitro high-throughput screening is vast, then the number of theoretical structures possible in virtual libraries is almost incomprehensible. A maximally diverse virtual library is neither practical nor possible, so attempts must be made to limit the size of the library. Compounds should be selected first on the basis of being synthesisable products: lengthy, expensive and low yield reactions would most likely not produce an economically viable product. Another method may be to restrict compounds to structures approaching those of current marketed drug products. The majority of commercial drugs can be represented by a limited number of structural scaffolds [Bemis & Murcko, 1996]. In addition, structures representing a smaller number of scaffolds only may be included. For example, only chemical structures known to be susceptible to a particular enzyme may be included. Three types of libraries can be defined in this manner: general, focussed, and targeted. In order of specificity, general libraries are the least specific and are designed to be arbitrarily of broad interest in high-throughput screening. Focussed libraries are aimed at a family of related targets, for example cytochrome P450 (CYP) 2C9 substrates, whereas targeted libraries are specific for one particular target [Walters et al., 1998].

Such libraries may also be screened for desirable structural characteristics. It is known that hydrogen bonding, lipophilicity, and molecular surface properties can
affect drug transport and membrane permeation [Stenberg et al., 2000]. These factors play an important role in bioavailability and, hence, suitability of a drug for manufacture as an oral formulation. Simple methods can be employed such as counting the number of hydrogen donors and acceptors, although more complex methods may be required for calculation of lipophilicity and molecular surface properties. The “Rule of Five” has been employed as a general guideline in industry to limit the size of virtual libraries to compounds likely to be adequately absorbed from the intestine [Lipinski et al., 1997]. The rule was developed upon examination of 2245 drugs from the World Drug Index (WDI) that were believed to have entered Phase II trials and were orally absorbed. According to the rule, compounds are deemed to have poor intestinal absorption if any two of the following conditions are met:

\[ \Rightarrow \text{There are more than five hydrogen-bond donors.} \]

\[ \Rightarrow \text{The calculated log } P\text{ (clog } P\text{) is greater than five.} \]

\[ \Rightarrow \text{The molecular weight (MW) is over 500.} \]

\[ \Rightarrow \text{There are more than 10 hydrogen-bond acceptors.} \]

Oxygen and nitrogen atoms are defined as being hydrogen bond acceptors, and –NH or –OH groups are defined as being hydrogen bond donors. Calculated log $P$ values may be determined using either a fragmental or molecular additivity approach depending on the nature of the data set [Lipinski et al., 1997]. The Rule of Five does not definitively categorise all well and poorly absorbed compounds, although it is simple, fast, and provides a reasonable degree of classification.
1.3 Modeling Techniques in Pharmacokinetics

Modeling provides a means to describe and understand data. It can also be useful for predictive purposes. Pharmacokinetic models are relatively simple mathematical tools that represent complex physiologic spaces or processes. Using a mathematical model such as a set of equations, large volumes of data may be summarised to provide a simpler representation. Depending on the modeling technique, pharmacokinetic data may be analysed using the model constructed. Thus, insight into mechanisms involved in pharmacokinetics such as distribution and elimination may be gained. Models which have been adequately validated can then be used for predictive purposes [Bourne, 1995]. This may be useful, amongst other things, in aiding lead compound selection or for failing unsuitable compounds early during the developmental processes.

Different modeling approaches can be more or less useful for a given modeling task. It is crucial, therefore, to select the most appropriate modeling technique for each situation. ANNs were chosen for the present research because of their demonstrated ability to develop predictive data-driven models. Other methods may be more useful if the aim is to, for example, focus on mechanistic relationships in pharmacokinetics. Nevertheless, the flexible nature of soft computing makes ANNs a potentially useful tool in predictive pharmacokinetic modeling. Such flexibility dictates that replicate experiments be performed to provide a measure of experimental precision.

A crucial aspect of modeling for predictive purposes is validation of the final model. Validation involves testing the ability of such a model to make predictions.
An unvalidated model is only useful for the data it was constructed on. Therefore, model testing using a cross-validation technique (Section 2.5.2) or independent testing compounds is often employed. Validation of a model provides a measure of its predictive ability and/or potential utility.

1.3.1 Non Structure-Based methods

The focus of the present research is on structure-based methods of modeling but non structure-based methods will be mentioned briefly here for completeness. Three approaches that have been suggested for pharmacokinetic modeling include the compartmental approach, physiologically based methods, and model-independent techniques.

The compartmental approach is an empirical approach which is based on a simple compartmental model. These compartments have no strict physiological or anatomical basis. The body is represented by a number of theoretical compartments that communicate reversibly with each other. The compartment can represent a body volume or, just as easily, it could represent a chemical state such as the metabolite of a drug. This approach usually uses either one or two compartments. Compartments are loosely considered a tissue or group of tissues with similar blood flow and drug affinity [Cutler, 1978]. Since there is more mathematical than physiological relevance for the parameters obtained in compartmental pharmacokinetic models, they cannot be used to extrapolate between species or provide mechanistic information about drug pharmacokinetics.

Despite its simplistic nature, many useful quantities can be derived using this approach and by comparing predicted values with actual data. They are also useful
when only plasma or blood concentration-time data are available without necessarily requiring tissue concentration data.

A physiologically based pharmacokinetic (PBPK) model identifies the compartments with actual body spaces. Such models are a great deal more complex than simple compartmental models. PBPK modeling incorporates physicochemical data as well as anatomical and physiological data from animals or humans to develop models for pharmacokinetic prediction [Grass & Sinko, 2002]. These models describe the mechanistic inter-relationships between ADME processes. Hence, they are more adaptable to clinical therapy and for changing situations. PBPK models can also be used for predictive purposes. One disadvantage is that they do require a large number of experimental parameters for model construction [Balant & Gex-Fabry, 2000]. For this reason, PBPK models are only rarely used in early drug development, although they can provide useful information regarding drug disposition and metabolism [Poulin & Theil, 2002].

Both compartmental and PBPK models require multiple data points from a single subject. In contrast, population pharmacokinetic modeling can use pooled data from multiple subjects. This is particularly useful when pharmacokinetic data is “sparse,” that is, when only limited data is available. In addition to sparse data, rich data can also be used for population pharmacokinetic modeling either separately, or in combination with sparse data [Tett et al., 1998]. One of the main advantages of population pharmacokinetic modeling is that data is gathered from a number of sources so fewer samples per subject are required. Another advantage is that pooled data allows conclusions regarding inter-subject variability in pharmacokinetics to be drawn. Disadvantages are that separate models must be
developed for each drug and models are only representative of the species in question. Therefore, population pharmacokinetic modeling is not often used in the early stages of drug development.

The model-independent approach is the most recent paradigm and is purely mathematical. It avoids recourse to kinetic parameters that may not be valid, and models developed tend to be less complex. This approach is good for modeling ADME values but gives no physiologically relevant information about drug properties.

**1.3.2 Structure-Based Methods**

The fundamental assumption in pharmacokinetic modeling based on structural considerations is that changes in molecular composition and atomic arrangement are quantitatively responsible for changes in drug pharmacokinetics. Such an assumption is based on the success of early QSAR models which demonstrated the relationship between pharmacologic activity and molecular structure [Seydel & Schaper, 1981]. Most of the reported activity values were from *in vitro* experiments using isolated organs or enzyme preparations. The challenge in moving from cellular systems to whole-body systems was, and still is, considerable. It has been proposed that structural variations may have a more obvious effect on pharmacokinetic parameters than pharmacological activity since they may be controlled to some extent by the physicochemical properties of a molecule. For example, the crossing of a biological membrane may be related to the lipophilicity of a molecule and, since the structure of biological membranes is consistent throughout the body, a consistent relationship between lipophilicity and
membrane permeability should be expected [Seydel & Schaper, 1981]. The complexity of living organisms dictates that no such simple correlation is apparent: although relationships can be approximated they may not entirely be explained by structure.

1.3.2.1 Multilinear Regression

The first QSAR models related physicochemical characteristics of a molecule with activity using mathematical regression equations. The Hansch analysis, or linear free-energy related (LFER) approach works on the principle that physicochemical properties of a compound are additive and may be combined linearly to approximate activity. This can be summarised in the following manner (Equation 1-1):

\[ y = \beta_1 x_1 + \beta_2 x_2 + K + \beta_k x_k + \epsilon \]  

*Equation 1-1*

where \( y \) is the independent variable and may be a biological activity or property, \( k \) is the number of independent variables, \( \beta_1 \ldots \beta_k \) are the regression coefficients, and \( \epsilon \) is a constant. The coefficients are determined by means of multiple linear regression using the least squares method. The assumption in LFER modeling is that the different magnitudes of a biological activity or property within a compound series correspond to changes in the free energy of the compounds which occur when reactions or interactions take place. The gradations of both activity/property and free energy are supposed to be linearly related [Hansch & Fujita, 1964]. Free energy is difficult to determine in biological systems so constants representative of free energy can be used. These constants include rate constants, and steric, electronic, and lipophilic parameters.
One of the advantages of modeling using such simple equations (Equation 1-1) is that direct relationships between variables and the target activity/property values are evident. A positive coefficient corresponds to an increase in the target value whilst the opposite is true for negative coefficients. The absolute size of the coefficient indicates the magnitude or importance of the contribution of a particular variable. Thus, if all data are scaled appropriately, a large absolute coefficient indicates an important contribution by a particular descriptor whereas the smaller the absolute value of a coefficient the less of an influence that descriptor has on the target value.

A disadvantage with multilinear regression is that in general about four or five compounds (patterns) at least are required for each variable used. This is a problem with small data sets since it may dictate that only a small number of descriptor variables can be used. This can then mean that insufficient information is available to enable construction of a meaningful model. Furthermore, the relationship between the number of patterns and number of variables needs to be monitored in multilinear regression to avoid chance effects [Topliss & Edwards, 1979]. Another disadvantage is that drug data often contains correlated or skewed information. This can then lead to the construction of poor regression models [Butina et al., 2002].

Early QSPkRs relied mostly on experimental variables to develop multilinear regression equations for a range of pharmacokinetic parameters. The most common variable used, log $P$, has been correlated with many different ADME parameters in both animals [Winningham & Stamey, 1970; Martin & Hansch, 1971; Seydel et al., 1980; Blakey et al., 1997] and humans (Table 1-1). Since the
present research is concerned with human QSPkR models, animal models will be discussed only briefly in relation to human models. Early QSPkRs were generally constructed using only small congeneric data sets. Experimental values were obtained for each study individually to ensure consistency of results. Animal data was much easier to obtain than human data so animal QSPkRs are more prevalent in the literature. Success of early QSPkR studies for prediction and drug development was limited due to the small number of drugs and types of descriptors used. In addition, models were developed to relate descriptor variables with pharmacokinetic parameters of drugs only present in the training set. Models were generally not further validated which limited their usefulness.

Table 1-1. Human QSPkRs using multilinear regression.

<table>
<thead>
<tr>
<th>Pharmacokinetic Parameter</th>
<th>Drug Data</th>
<th>Equation Variables</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Half life</td>
<td>sulfonamides</td>
<td>log $P$</td>
<td>[Seydel et al., 1973]</td>
</tr>
<tr>
<td>Epidermal absorption</td>
<td>aliphatic alcohols</td>
<td>log $P$</td>
<td>[Lien, 1975]</td>
</tr>
<tr>
<td>Protein binding</td>
<td>penicillins</td>
<td>log $P$, log $P$</td>
<td>[Craig &amp; Welling, 1977]</td>
</tr>
<tr>
<td>Volume of distribution</td>
<td>penicillins</td>
<td>$P$</td>
<td>[Watanabe &amp; Kozaki, 1978a; Watanabe &amp; Kozaki, 1978b]</td>
</tr>
<tr>
<td>Metabolism by monoamineoxidase</td>
<td>aliphatic amines and alcohols</td>
<td>log $P$, pKa</td>
<td>[Kubunyi, 1979]</td>
</tr>
<tr>
<td>Clearances, mean residence time, volumes of distribution</td>
<td>β-adrenoceptor antagonists</td>
<td>p$K_a$, $K_{SF}$ (octanol-buffer partition coefficient)</td>
<td>[Hinderling et al., 1984a; Hinderling et al., 1984b]</td>
</tr>
<tr>
<td>Clearances, mean residence times, volumes of distribution</td>
<td>non-congeneric compounds (17)</td>
<td>MW, intrinsic and alcohol solubilities, protein binding, distribution coefficient</td>
<td>[Herman &amp; Veng-Pedersen, 1994]</td>
</tr>
</tbody>
</table>
While obviously useful, log $P$ does not provide all the information about a molecule necessary for construction of unlimited sorts of structure-property relationships. Additional physicochemical descriptors have been incorporated in models over time and eventually theoretical descriptors were included as well [Genty et al., 2001]. Even so, the limitations of multilinear regression were still apparent: numbers of descriptors needed to be controlled and validation of models remained a challenge.

1.3.2.2 Artificial Neural Networks

In comparison with multilinear regression, ANNs are more flexible, robust, and better at prediction [Butina et al., 2002]. They are inherently nonlinear in nature so are better able to manage the complex and nonlinear data as is often present in the pharmaceutical field. Furthermore, they have demonstrated robustness with noisy experimental data [Turner et al., 2003]. Their introduction into the area of pharmacokinetics has been relatively late compared with other scientific and industrial fields. ANNs have found use in clinical monitoring to match pharmacokinetic profiles with pharmacodynamic effects [Minor & Namini, 1996], and for predicting patient creatinine clearance based on physiological variables [Herman et al., 1999]. In both studies, ANNs produced results superior to regression or other modeling methods alone, and conclusions were that ANNs would be of particular benefit in those clinical situations. Similarly, another study used physiological and demographic data to predict the pharmacokinetics and pharmacodynamics of repaglinide, an oral hypoglycaemic agent [Haidar et al., 2002]. In addition to obtaining acceptable predictions, the ANN technique also allowed identification of significant covariates. It is important to note that ANNs
do not present mechanistic information so other methods must be used should the mechanisms underlying pharmacokinetics be required. Even so, similar predictive models have been obtained using ANNs when compared with mechanistic modeling techniques [Nestorov et al., 1999].

ANNs have also been compared with conventional programs for construction of predictive population pharmacokinetic models. In comparison with the population pharmacokinetic modeling program NONMEM [GloboMax LLC, 1998], ANNs demonstrated lower absolute and prediction errors [Chow et al., 1997]. Modeling of the pharmacokinetic data was accomplished well by both methods, however, results presented were not validated against external data. A more recent study trained both ANNs and NONMEM on two thirds of a 622 point data set and found that prediction of the remaining one third of the data was consistently superior using the ANN model [Tolle et al., 2000]. In both these studies, the same covariates were used for ANN and NONMEM model construction. Hence, there appears to be good potential for ANN application in population pharmacokinetic modeling.

QSAR studies have benefited from the use of ANNs for over a decade [Aoyama et al., 1990]. Following the progress of QSAR, development of QSPkRs using ANNs has also advanced. Both physicochemical and theoretical descriptors have been used successfully in ANN QSPkR studies, either individually or in combination [Ritschel et al., 1995]. The trend has been towards completely in silico models as described earlier, since the speed associated with ANN models coupled with cheaper and more powerful computational methods has made this
direction more feasible. A more detailed discussion of the application of ANNs in QSPkR modeling is given in Section 1.6.

1.4 Artificial Intelligence Systems

Soft computing methods have been used to varying extents in pharmaceutical research. ANNs have been the most popular due their intrinsic nonlinearity and characteristic robustness. They have been termed “universal approximators” since by varying network architecture it is possible to model almost any given situation [Haykin, 1994]. In addition to simply modeling a system, their use in prediction has also received growing attention. Multilayer perceptron (MLP) and radial-basis function (RBF) ANNs have been used more extensively than other ANN paradigms, and they will be discussed in the proceeding sections. The Kohonen topology-preserving map, otherwise known as self-organising map (SOM) is a form of ANN which may be used as a clustering tool. In theory, relationships in a multidimensional space are mapped onto the surface of a torus so that the Euclidean distance separating each point is equal. Sectioning one part of the torus and unbending it produces a cylinder. Sectioning the cylinder along its length allows it to be unrolled to form a rectangle. Hence, the original points now reside on a 2-dimensional (2D) plane. One advantage of self-organising maps is that target values are not required for initial model construction. Genetic algorithms (GAs) are evolutionary systems based on chromosomal recombination and selection. They are often used in combination with other soft computing methods such as feed-forward back-propagation ANNs to select optimal subsets of descriptors in QSAR/QSPR studies [Terfloth & Gasteiger, 2001].


1.4.1 Multilayer Perceptron ANNs

ANNs are mathematical models based on the structure of the biological brain, being composed of many individual processing units or artificial neurons which are extensively inter-connected to form a network. Emulation of brain function was based on the hypothesis that the information in a brain resides in the strength of connections between neurons and not in the internal state of the neurons themselves [Bucinski et al., 2000]. Learning is simply the adjustment of the strengths associated with each connection. For this reason, connections between neurons are known as “weights.” The neurons are connected to one another in parallel which provides their characteristic speed, robustness, and generalisation ability (Maddalena, 1998).

Multilayer perceptrons are of the feed-forward back-propagation class. They are composed of neurons organised into an input layer, one or more hidden layers, and an output layer (Figure 1-2). The number of input neurons is equal to the number of variables, and the number of output neurons is equal to the number of targets being predicted which in most cases is equal to one. The number of neurons in the hidden layer can be either one, for studies which parallel multilinear regression, or greater than one for studies dealing with nonlinear data [Aoyama & Ichikawa, 1991]. In each layer there is usually also an extra bias neuron which is not connected to the neurons in the previous layer. The bias neuron provides a magnitude adjustment to the input values so that they are in the correct range for processing by other neurons [Swingler, 1996]. Since the bias neuron is connected to each and every neuron in the hidden and output layers, the relationship between input variables and target output is not easily traceable. As
mentioned previously, in multilinear regression the influence of each variable is proportional to the size of its coefficient whereas this is not necessarily the case with multilayer perceptrons.

\[ \text{Input layer} \]
\[ \text{Hidden layer} \]
\[ \text{Output layer} \]

**Figure 1-2.** Schematic representation of a multilayer perceptron ANN.

In order for the multilayer perceptron to make predictions it must first be trained as described in the General Methodology (Section 2.5.1.1).

Multilayer perceptron ANNs have been examined for use in a number of areas in pharmacy. In a clinical setting, ANNs were used to monitor the pharmacodynamics of short-acting neuromuscular blockers [Lendl et al., 1999]. ANNs were chosen since they offered a fast and controllable mechanism for prediction without the need for more costly biopharmaceutical data. Compared with conventional closed-loop controllers, results using the ANN were encouraging, suggesting a potential use for this technique in this and other clinical settings. In another clinical study both pharmacokinetic and pharmacodynamic relationships were analysed using ANNs [Minor & Namini, 1996]. It was also
suggested that ANNs may be useful for time-dependent modeling and aiding in the development and analysis of clinical trials.

Immunosuppressant therapy in organ transplant recipients requires close monitoring of drug concentrations to ensure adequate immunosuppression. Conventional methods rely on measurements of trough blood concentrations, although free plasma concentrations and two hour post-dose concentrations are also being examined. Using population data, clinical monitoring of peak and trough serum concentrations of gentamicin was examined [Brier & Aronoff, 1996]. Prediction of peak concentrations using ANNs was comparable with models constructed using NONMEM, while prediction of trough concentrations was superior using ANNs.

In other predictive applications the scaling-up of allometric data to predict human pharmacokinetic parameters has been performed with ANNs. In one study, animal data taken from the literature was used to predict human volume of distribution and clearance values [Hussain et al., 1993]. Although ANNs were shown to provide acceptable models, problems included the requirement of a substantial amount of training data which may not always be readily accessible. However, it was also shown that existing animal data was able to be supplemented with theoretical data from drug structure. Furthermore, drug physicochemical data may also be included for construction of pharmacokinetic models [Ritschel et al., 1995]. Prediction of clearance and volume of distribution using ANNs with such information was shown to be similar to in vitro estimations, so no superiority of either technique was apparent in that respect. However, time and cost savings gained using the ANN indicated that it was potentially a more useful technique.
As with the majority of earlier studies using theoretical descriptors, little attempt was made to explain the relationship of such descriptors with the pharmacokinetic parameters in question.

1.4.2 Radial-Basis Function ANNs

Radial-basis function ANNs differ from multilayer perceptrons in that the nonlinear transformation of data occurs only in the hidden layer and not elsewhere [Yao et al., 2002b]. Radial-basis function ANNs belong to the class of kernel estimation methods and employ a transfer function representing a bell-shaped Gaussian response surface. In contrast, the transfer function in multilayer perceptrons is generally sigmoidal (Figure 1-3). Although they only have three layers of neurons available, radial-basis function ANNs are functionally similar to multilayer perceptrons. Details of training are given in General Methodology (Section 2.5.1.2).

![Figure 1-3. Transfer function for a) hidden neurons in a radial-basis function ANN, and b) neurons in a multilayer perceptron.](image)
Radial-basis function ANNs have not been used as extensively as multilayer perceptrons in pharmacokinetics or elsewhere. For the same modeling task they generally require more neurons in the hidden layer than the latter which leads to increased network complexity. It has been suggested that complexity of a network can influence the training and predictive performance of a model. The analogy applied is that multilinear regression studies require a certain minimum number of patterns per optimisable parameter. That is true for any modeling technique, however, the analogy is not entirely applicable to ANNs in general since they are nonlinear systems. Hence, the relationship between the complexity of the model and the number of patterns depends specifically on the nature of the model itself [Turner et al., 2003].

One advantage of radial-basis function ANNs over multilayer perceptron ANNs is the speed at which they are trained. In one QSPR study using 233 compounds both paradigms were directly compared [Tetteh et al., 1996]. It was found that results using both paradigms were similar for both training and validation. However, radial-basis function ANNs trained faster and were less likely to fall into local minima than multilayer perceptrons that employed sigmoidal transfer functions. Further comparisons were also made with linear modeling techniques which were found not to be as useful as ANN models. A similar conclusion was also drawn in other comparative studies, for example, in a recent QSPR for benzene derivatives employing quantum mechanical values as input descriptors [Wang et al., 2002].

Since radial-basis function ANNs train relatively fast, they are well suited to problems involving large data sets with numerous descriptor variables. Several
studies have generated a number of theoretical descriptors and then selected only a subset to use in the final model. One such study generated 35 topological descriptors and selected a subset of 9 for the final QSPR of 173 compounds [Yao et al., 2002a]. Similarly, another QSPK for intestinal permeability constructed at a predictive model of 15 descriptor variables from a total of 57 generated for 86 drugs [Agatonovic-Kustrin et al., 2001]. Both studies employed different methods for selection of optimum descriptors, but all models were suitably cross-validated to ensure soundness of results. The fact that radial-basis function ANNs require a greater number of hidden neurons than multilayer perceptrons does not usually affect training time significantly.

1.4.3 Other Soft Computing Methods

Genetic algorithms are an evolutionary technique well suited for selection purposes. As such, they have been used to reduce the number of available compounds [Gobbi & Poppinger, 1998] and for compound selection optimisation to identify bioactive molecules [Gillet et al., 1998] in virtual libraries. As well as being a useful tool in isolation, genetic algorithms have often been applied in combination with other modeling techniques such as ANNs. These genetic neural networks (GNNs) provide an automated pruning technique for studies involving large numbers of descriptor variables [Zupan & Novic, 1999]. Comparatively, GNNs have matched QSAR results obtained using manual pruning, and have the added advantages of speed and unbiased descriptor selection [So & Karplus, 1996]. With the increasing number of theoretical descriptors able to be generated from drug structure, GNNs have also successfully aided the selection of key descriptors for QSPR models constructed using multilinear regression [Turner et
al., 1998] and ANN [Agatonovic-Kustrin et al., 2001] methods. Although genetic algorithms present a useful alternative to manual selection of descriptors, they tend not to be used for exhaustive searching or correlating since they are computationally expensive relative to other in silico approaches.

A Kohonen self-organising map represents an unsupervised neural network paradigm, and is essentially a 2D representation of a multi-dimensional space [Kohonen, 1997]. With respect to pharmaceutical applications, a drug may be described by numerous physicochemical or theoretical descriptors and then represented by a position on a 2D map relative to other compounds. Thus, similar compounds are clustered together on the self-organising map whilst different compounds are positioned away from each other. This technique has been used to classify pharmacologically active molecules amongst non-congeneric data sets [Bauknecht et al., 1996] and also to locate potentially useful anticancer drugs [van Osdol et al., 2000]. Clustering of compounds according to odour properties has also been performed [Audouze et al., 2000] but it was found that appropriate description of the compounds for projection onto a self-organising map was problematical. Useful self-organising maps have been created in other areas such as protein surface [Stahl et al., 2000] and molecular surface potential mapping [Gasteiger et al., 1994]. However their utility in pharmaceutical product development is limited.

Fuzzy logic is characterised by a mathematical framework that lacks well-defined boundaries. The lack of strict boundary conditions allows flexibility in classification problems, and fuzzy sets have successfully been developed to classify compounds according to their chemical composition [Pop et al., 1996],
and similarity [Maggiora & Mezey, 1999]. In pharmaceutics, fuzzy sets have been applied in clinical drug dosage monitoring [Kern et al., 1997; Shieh et al., 2002], and prediction of serum pharmacokinetics [Sproule et al., 1997]. When compared with conventional population pharmacokinetic modeling using the same data, fuzzy logic provided comparable results and allowed determination of important covariates. Other pharmacokinetic studies have used fuzzy methods to predict human bioavailability and volume of distribution [Hirono et al., 1994a; Hirono et al., 1994b]. These studies grouped compounds according to known bioavailability values and further subdivided groups according to broad chemical composition. Even though fuzzy logic appears to be a useful alternative to conventional modeling, further development is required to enable prediction of unknown compounds.

1.4.4 Descriptor Selection

Optimum descriptor selection remains a fundamental problem in QSAR/QSPR studies. Many descriptors may be considered for inclusion in a model, however not all may provide useful information and indeed some may even be detrimental. The aim of descriptor selection is to improve model generalisation by reducing unnecessary data [Tetko et al., 1998].

Early multilinear regression models utilised smaller numbers of physicochemical descriptor variables such as log $P$ and association constants. As the field of QSAR/QSPR grew, additional ways of describing compounds were realised and methods had to be established to select relevant descriptors. Since multilinear regression provides a direct relationship between a descriptor and the output
Chapter 1. Introduction

space, normalised descriptors with very low coefficients are considered not to contribute significantly to a given model and so may be removed without much harm. More complicated stepwise regression techniques have also been implemented. These involve identification of an initial model and then repeated alteration of the model from the previous step by the addition (forward stepwise) or removal (backward stepwise) of a descriptor variable. The search is terminated when stepping does not further improve the model.

Partial least squares regression is an extension of multilinear regression, but derives factors from the descriptor variables to maximise the covariance between the descriptor and output spaces [Bjork & Danielsson, 2002].

Clustering methods are also available which group similar descriptors together to minimise the variance within clusters but maximise variance between clusters. From these clusters, suitable descriptors may then be selected which should represent a substantial portion of the information contained in the entire descriptor set.

Principal component analysis also derives descriptors representative of the whole descriptor space but does so by linearly combining variables to maximise variance between the individual principal components [Bruni et al., 2002]. All these descriptor selection methods have been used to some extent in pharmaceutics [Abuzaruraloul et al., 1998] and QSAR, although other genetic algorithm and ANN techniques have proven more effective [Winkler et al., 1998].

As described earlier, genetic algorithms provide an effective means of descriptor selection which may then be combined with, for example, ANN modeling [So &
Karplus, 1996]. Chromosomes, representing the entire descriptor space, are composed of genes, which represent individual descriptors, and randomly crossed-over to simulate biological evolution. A fitness function applied to the resultant offspring retains the better-performing chromosomes, and then the process is reiterated until the chromosome with the best genetic composition has evolved [Zupan & Novic, 1999]. Analogous to biological systems, mutations are sometimes incorporated to help offspring avoid local minima. One limitation with genetic algorithm searching is that chromosomes are often constrained to a fixed length, which may restrict the characteristics of the terminal offspring. Variable-length chromosomes have been used, however, this attenuates the problem of lengthy training times [Yasri & Hartsough, 2001].

In contrast, studies using ANNs alone are computationally inexpensive. Such studies often include the entire descriptor set initially followed by descriptor selection being performed on a continuous basis until an optimum subset is achieved. Various selection or pruning techniques exist which aim to eliminate redundant weights and/or descriptors, thus leaving only those offering a significant contribution to the model.

Pruning may be divided into sensitivity and penalty term methods, and can be implemented either manually or incorporated within the training algorithms. Sensitivity-based methods have been examined and it has been found that simpler magnitude-based algorithms performed as well as more sophisticated error-based algorithms [Tetko et al., 1996]. Penalty terms were also shown to be useful in removing redundant weights and were thus able to accentuate the importance of certain descriptors with respect to the target output space. Manual selective
pruning has been used successfully to reduce the number of descriptors and, although time-consuming, has the advantage of allowing greater control over the pruning process than automatic algorithm-based techniques [Maddalena & Johnston, 1995].

Details of the descriptor selection techniques used in this thesis are given in the General Methodology, Section 2.4.

1.5 Descriptors Used in Modeling

Classical physicochemical descriptors such as log $P$ are not available for all known chemical entities. Conversely, theoretical descriptors may be calculated for all chemical entities should the structure be known [Devillers, 1999]. Descriptors may be a scalar representation such as atom-counts, or rely on a matrix, for example topological indices, or require lattice-type information to allow calculation of 3D descriptors. There are currently over a thousand theoretical descriptors that have been applied to chemical- and drug-related problems. Many theoretical descriptors contain similar information and this is particularly true for descriptors derived for smaller molecules or for structurally similar compounds. It is generally accepted that appropriate methods should be undertaken to limit the number of topological indices in a study to those containing independent and useful information [Basak et al., 2000a]. Hence, suitable clustering or pruning is required to ensure that redundant descriptors are not included in the modeling process.
Independent descriptors are those which are not significantly linearly correlated with one another, thus, their information content is independent of that of other descriptors. To maintain diversity of information highly correlated descriptors are often excluded from a model. Even so, correlated descriptors may still be included in a successful models since, unless identical, they all contain a certain amount of independent and possibly useful information [Consonni et al., 2002].

Some of the more important theoretical descriptors and those relevant to the current project will be described in the proceeding sections.

1.5.1 Constitutional Descriptors

Constitutional descriptors are the simplest of all theoretical descriptors, although they are not strictly classed as topological indices. They encode basic information such as the number and type of atoms in a molecule, and they also include counts of functional groups.

There are two ways of presenting constitutional descriptors: the first is using a binary system where the presence or absence of a particular moiety is denoted by a one or zero respectively. The second and more common method quantifies the number of cases of each moiety. For example, bepridil is represented in a different manner according to each system (Table 1-2). The assumption behind constitutional descriptors is that variations in atomic and functional group composition influence whole molecule properties. Indeed, this assumption underpins QSAR studies for drugs defined by a common template structure and altered at various substituent positions around that template [Maddalena & Johnston, 1995].
The fact that constitutional descriptors are easily interpreted is a distinct advantage over more complex topological indices. Constitutional descriptors have been used for prediction of physicochemical parameters [Burden, 1996] through to entire QSPkR analyses in combination with other descriptors [Agatonovic-Kustrin et al., 2001]. For prediction of biological activity, the ability of constitutional descriptors to encode useful information appears to be better suited to congeneric series of compounds since deviations in activity can be directly attributed to variations in substituents [Jaen-Oltra et al., 2000].

Application to non-congeneric series of compounds has been performed, although, to account for the larger differences in molecular structure, usually more complex descriptors are required as well. These additional descriptors may also be constitutionally-based such as the $V3$ and $V4$ indices which denote vertices of valence three and four respectively, and $L$, which is defined as the topological length of the two most separate points on the graph [Galvez et al., 1994b]. Other constitutional descriptors account for features such as number of rotatable bonds.
and molecular mass derivatives, and these have been included in models of drug solubility characteristics [Jorgensen & Duffy, 2002] and other pharmacokinetic parameters [Herman & Veng-Pedersen, 1994] with reasonable success.

### 1.5.2 Topological Indices

Topological indices mathematically encode information regarding the structure of molecules which have been depicted as graphs. The molecular graph is comprised of vertices which correspond to atoms and edges corresponding to the bonds between these atoms. Often they are sensitive to size, shape, branching, cyclicity and, to a certain extent, electronic characteristics of molecules [Todeschini & Consonni, 2000]. Subgraphs are defined as two or more vertices connected by a bond or common path. Subgraphs may include branched and cyclic structures, and can have up to as many vertices as the entire molecular graph.

The seminal contribution to the field of topological indices was the introduction of the Wiener index. The Wiener index is defined as the sum over all bonds of the product of the number of vertices on each side of the bond [Wiener, 1947]. This index has been used extensively in the construction of QSPRs and QSARs for structurally related [Zakarya et al., 1993] and unrelated [Galvez et al., 1995] drugs. Performance has also been improved following modification of the single Weiner number to extended Weiner indices [Estrada, 1999].

The next significant topological index to be developed was the branching index proposed for a series of alkanes [Randic, 1975]. The Randic branching index, a precursor of the Kier and Hall connectivity indices, has led to successful elucidation of numerous topological-based QSARs and QSPRs. However, the
physicochemical significance of the Randic index was undefined for decades after its inception. Only recently has the link between these theoretical numbers and their relation to physical chemistry been revealed. Research demonstrating that the branching corresponds to the relative area of accessibility of a molecule has established that physical meaning can be extracted from theoretical descriptors [Estrada, 2002a].

### 1.5.2.1 Connectivity Indices

Kier and Hall connectivity indices, also called chi ($\chi$) indices, were developed to calculate zero- and higher-order connectivity descriptors [Kier & Hall, 1977]. Numerous correlations between connectivity indices and both physicochemical properties [Reinhard & Drefahl, 1999] and biological activity [Kier & Hall, 1986] of drugs have been identified, mostly for structurally related compounds.

**Table 1-3.** Information content of Kier and Hall connectivity indices [Kier, 1987].

<table>
<thead>
<tr>
<th>Index</th>
<th>Information Content</th>
</tr>
</thead>
<tbody>
<tr>
<td>$0\chi$</td>
<td>General features about atoms or points, including molecular volume, molar refractivity, density and magnetic susceptibility</td>
</tr>
<tr>
<td>$1\chi$</td>
<td>Number of atoms in the molecule, and related surface area and volumes, relative branching in structural isomers</td>
</tr>
<tr>
<td>$1\chi^v$</td>
<td>Molar refractivity, orbital electronegativity, molecular polarity, structural differences for six-membered rings</td>
</tr>
<tr>
<td>$2\chi$</td>
<td>Information about branching (three-atom fragments)</td>
</tr>
<tr>
<td>$3\chi_p$</td>
<td>Flexibility, conformational gauche-anti rearrangements</td>
</tr>
<tr>
<td>$3\chi_c$</td>
<td>Branching, density, multiplicity of “cross-road” atoms</td>
</tr>
<tr>
<td>$4\chi_{pc}$</td>
<td>Structural description of substituted aromatic rings and information about the orientation of ring substituents</td>
</tr>
<tr>
<td>$4\chi_{vc}$</td>
<td>Number of benzene ring substituents, the substitution pattern, length of the substituents up to three bond lengths, and heteroatom type of substituent (in conjunction with $4\chi_{pc}$)</td>
</tr>
</tbody>
</table>
These descriptors are theoretical in nature so the absolute meaning of each index is not easily described. However, since the method of calculation is well defined then certain indices may describe some specific features of a molecule (Table 1-3). Moreover, recent work for both structurally related and structurally unrelated compounds has revealed the relationship between connectivity indices and molecular accessibility area [Estrada, 2002b]. Since molecular accessibility area is important in chemical interactions then such a relationship demonstrates the relevance of topological indices in drug models. Connectivity indices mathematically describe molecular structure by encoding branching and cyclicity (nonvalence $\chi$) and heteroatom influence (valence $\chi$). They cannot, however, encode absolutely every single structural feature of a molecule. For example, structures exhibiting cis/trans isomerism and atomic chirality are not differentiated from one another by connectivity indices [Cao & Yuan, 2002]. Modification of connectivity indices has been proposed to account for such shortcomings [Basak et al., 2000c], although simple connectivity indices still remain popular in the broader scientific and industrial community. Details of the methods of calculation of connectivity indices are given in Section 4.2.2.1.

The application of simple connectivity indices extends from lead compound searching [Casaban-Ros et al., 1999] to QSARs and QSPkRs [Cercos del Pozo et al., 1996]. Studies have been performed for structurally diverse drugs and individual indexes have been shown to be important for different activity and property parameters. Linear combination of connectivity indices, for example differences and quotients, can describe features such as number and nature of heteroatoms, as well as inductive and mesomeric effects of molecules [Galvez et
al., 1994b]. Hence, they provide a valuable adjunct to the information presented by connectivity alone and have been utilised successfully in structure-pharmacokinetic studies [Rose et al., 2002].

1.5.2.2 Electrotopological Indices

Biological and chemical properties of molecules rely on both their structural and electronic attributes. Since topological indices such as the Wiener index and connectivity indices describe features of a molecule principally from a structural perspective, topological charge indices were developed to explicitly describe the charge distribution characteristics in a molecule.

The topological charge indices, $G_k$, encode the total charge transfer between atoms in a molecule at a distance $k$ from one another [Galvez et al., 1994a]. Thus, $G_k$ indices are related to the dipole moment of a molecule, and can be of the order one to $L$. For acyclic compounds, $J_k$ indices represent the mean value for the charge transfer across the molecule and are a modification of the corresponding $G_k$ indices. Details regarding calculation are given in Section 4.2.2.1. Although originally defined for acyclic alkanes, $G_k$ and $J_k$ have been used to model physicochemical properties and biological activity of both structurally similar and structurally diverse drugs, including cyclic compounds [Galvez et al., 1994b; Galvez et al., 1995].

Electrotopological state indices, $S_n$, are based on the “intrinsic state” of atoms which is related to their valence state [Kier & Hall, 1990]. Intrinsic states have been defined for 39 different atom valence states, which allow calculation for a wide range of molecular structures. In addition to structure-activity relationships,
these electrotopological state or E-state indices have successfully been correlated with aqueous solubility [Huuskonen et al., 1998] and blood-brain partitioning in combination with connectivity indices [Rose et al., 2002]. From this work E-state indices appear to have potential for use in QSPkR modeling.

1.5.3 Quantum Chemical Numbers

Typical quantum chemical numbers include energies of the lowest unoccupied molecular orbital (LUMO) and highest occupied molecular orbital (HOMO), dipole moment, dielectric energy, steric energy, total energy, minimum energy, heat of formation, and electron affinity. Quantum chemical calculations rely on the 3D structure of molecules. Descriptors obtained in this manner are therefore sensitive to conformational changes of a compound. Molecules typically undergo an energy minimisation routine \textit{in silico} in order to obtain the anticipated \textit{in vivo} 3D conformation. Even though quantum chemical numbers provide information representing absolute thermodynamic and electronic properties for a molecule, these values may not apply to an \textit{in vivo} situation where the conformation of the molecule differs from the proposed \textit{in silico} representation. Thus, caution should be exercised when interpreting the meaning of quantum chemical numbers at a detailed level. They have, however, provided useful information on a broader scale from complexation [Estrada et al., 2001] to structure-pharmacokinetic studies [Ekins & Obach, 2000]. Moreover, it is easier to interpret the physicochemical meaning of quantum chemical numbers than it is to interpret other topological indices.
Several studies have demonstrated the importance of quantum chemical considerations in the intestinal permeability of structurally diverse compounds. One study examined 18 theoretical and quantum chemical descriptors using principal component analysis [Winiwarter et al., 1998]. It was established that the information content in those principal components was sufficient to indicate a relationship between structure and permeability. It was found, however, that quantum chemical numbers did not rank as highly as other theoretical partitioning and solubility descriptors, and that better models were constructed without the quantum chemical numbers.

Another study generated 42 theoretical descriptors for 254 drugs, and reduced the number to 12 in the final model. Of those 12, five were quantum chemical numbers, and their influence on membrane penetration was quantified [Agatonovic-Kustrin et al., 2001]. It was found that the significance of dielectric energy was more than double the next most important descriptor. Alone, quantum chemical numbers have been correlated with metabolism properties in rats [Cupid et al., 1999]. The urinary excretion of a series 22 of benzoic acid analogues and their metabolites was modeled using linear regression. Several models were developed and reasonable prediction correlations were achieved. Owing to the complexity of metabolic pathways, development of structure-metabolism relationships is considerably difficult. It has been demonstrated that relationships can be developed for structurally similar compounds. Therefore, the next challenge would be to extend models to include large numbers of structurally diverse compounds.
1.5.4 Solubility and Partitioning

As described earlier, solubility and lipophilicity are vital elements in determining the entry of drugs into the body via the oral route. Solubility characteristics can limit absorption from the GI tract while oil/water partitioning can affect drug distribution and binding to proteins. Many studies have correlated physicochemical as well as theoretical descriptors with experimental solubility and oil/water partitioning, and this is dealt with in Section 1.6.1. Calculated solubility and solubility-related parameters are often related to the ionic and electronic characteristics of a molecule. One scheme represents solubility as a combination of dispersion, polarity and hydrogen bonding values to give a vector in 3D space which describes a “radius of interaction” of a molecule [Hansen, 1967]. In contrast, calculation of partition coefficients is performed according to an additivity method. One approach for generation of calculated log \( P \) (clog \( P \)) sums the contribution to lipophilicity at an atomic level [Viswanadhan et al., 1989], while another employs contribution of functional groups to lipophilicity [Hansch, 1979]. There have been improvements suggested for both of these methods [Wildman & Crippen, 1999]. Nevertheless, clog \( P \) values obtained using either method have been validated using large numbers of compounds, and have proven useful in a broad range of structure-activity/-property relationship applications [Lipinski et al., 2001].

Understandably, studies examining intestinal permeability have found clog \( P \) to be an important descriptor. One study developed a number of models using different combinations of descriptors and found clog \( P \) to have large regression coefficients in several models [Winiwarter et al., 1998]. Similarly, an ANN study
using numerous theoretical descriptors determined clog $P$ to have the greatest effect on the model [Agatonovic-Kustrin et al., 2001]. The importance of clog $P$ as an indicator of bioavailability is also apparent from its inclusion in the Rule of Five [Lipinski et al., 1997] which is utilised widely in the drug development industry.

### 1.5.5 Other Descriptors

A multitude of other descriptors exist, some with easily identifiable meaning and others more abstract. Geometrical and bulk descriptors provide information regarding the 3D characteristics of a molecule. Calculated surface area, molar volume, and solvent accessible area all depend on the particular conformation adopted by a molecule \textit{in silico}. Hence, a suitable approach to determine the 3D conformation of molecules must be employed to ensure validity of descriptors. These descriptors have not been used to a great extent in the literature for QSPkR analyses due to their dependence on molecular conformation and their lack of relevance to QSAR studies. QSPkRs rely on the ability of drug molecules to be absorbed and excreted which in turn depends on the size and shape of molecules. For example, size and shape characteristics can affect glomerular filtration, and they have been shown to affect the rate of membrane permeability [Ghafourian & Fooladi, 2001]. Accordingly, geometric and bulk parameters are potentially more useful for QSPkRs than QSARs.

Many topological descriptors rely in a specific representation of a molecule as a graph or matrix. By changing the manner of representation, different descriptors encode for diverse theoretical characteristics of a molecule. It would be prudent,
therefore, to include numerous theoretical descriptors in the initial stages of model construction to adequately represent the multidimensional nature of a molecule.

1.6 Structure-Pharmacokinetic Relationships

The ANN modeling technique utilised in the present research has been employed extensively in QSAR analyses over the last decade. There are many examples of activity, toxicity, and carcinogenicity studies in the literature, however, the reader is directed to the following references [Maddalena, 1996; Basak et al., 2000b; Buchwald & Bodor, 2002; Greene, 2002] to avoid unnecessary discussion in this thesis. Other structure-property applications in pharmaceutics include formulation optimisation [Takayama et al., 1999], partition coefficient prediction [Huuskonen et al., 2000b], infrared spectra analysis, and chromatographic retention modeling [Agatonovic-Kustrin & Beresford, 2000].

Early QSPkR model development relied on regression equations to correlate physicochemical properties of drug molecules with pharmacokinetics [Haj-Yehia & Bialer, 1989]. It was suggested that QSPkRs should be confined to structurally similar compounds in order to avoid the risk of encountering discontinuities in pharmacokinetic properties. For example, log $P$ for a set of sulfonamides can determine plasma protein binding but may not effectively represent protein binding of penicillins. It was also proposed that models explaining only 60% of the variance for a particular pharmacokinetic parameter could be deemed adequate for the purpose of providing useful information about a particular congenic series of drugs [Seydel & Schaper, 1981]. Considering the high stakes resting on
successful drugs reaching the market as well as the cost of development of apparent failures, such a poor figure may now no longer be acceptable.

Recently, more functional QSPkRs have been developed to model the pharmacokinetics of structurally diverse sets of drug data [Herman & Veng-Pedersen, 1994]. Although still employing linear regression techniques, these QSPkR studies demonstrated that prediction of pharmacokinetic parameters was not necessarily limited to congenic series of compounds. The use of physicochemical descriptors was the first logical step due to their general acceptance in other QSAR and QSPR studies. Theoretical descriptors have also been considered in combination with physicochemical descriptors to account for diffusion characteristics which may influence pharmacokinetics [Herman & Veng-Pedersen, 1994]. Finally, QSPkRs constructed solely from theoretical descriptors have shown promise in locating biologically active compounds amid structurally diverse drugs and also in modeling distribution half life [Galvez et al., 1996]. The regression techniques employed allowed simple models to be constructed. However, it has been the use of more robust soft computing methods that has increased over time instead.

1.6.1 Absorption

Absorption at the site of administration can influence drug bioavailability. Drugs administered intravenously do not undergo absorption processes, whereas other routes of administration generally require absorption to occur before a drug is available to the body. For orally-delivered formulations in particular, dissolution, absorption, and first-pass metabolism all contribute to the bioavailability of a
drug. For a drug to be absorbed it must first go into solution in order to cross biological membranes (Figure 1-4). For an orally-delivered drug that has poor dissolution characteristics the time spent at absorption sites in the GI tract may be insufficient for complete absorption to occur. In such a case residence time in the GI tract can be increased by slowing intestinal motility, however, bioavailability may still be limited by the drug solubility.

![Figure 1-4. Sequence of events in solid oral drug absorption.](image)

One study examining three different data sets containing congeneric drug compounds aimed to predict aqueous solubility using topological indices [Huuskonen et al., 1997]. Solubility data was taken from the literature and cluster analysis was used to select an uncorrelated subset of five descriptors from the many descriptors that were generated. Models were subjected to the leave-one-out (LOO) cross-validation testing to overcome over-training and also as an indication of predictive ability. Even though it does not provide a true measure of the predictive ability of a model LOO cross-validation is useful when the size of data sets is limited.

The results obtained demonstrated a number of important points. First, successful models were able to be constructed using simple calculated descriptors rather than from experimental data. Second, ANN models were found to be more robust than regression models of similar data and, third, some models required more than a single class of topological indices to enable reasonable prediction. In further developments, other topological indices have been used to construct models for
drugs not part of a congeneric series. Since 3D structure is important in dissolution, descriptors encoding geometrical properties of a molecule as well as charge distribution have been correlated with aqueous solubility using both regression and ANN techniques [Bodor et al., 1991]. Since 3D information was required, energy minimisation routines were applied to drug structures to arrive at appropriate conformations. True predictive ability was tested with an independent set of compounds and in most cases ANN models were found to be superior to regression models. The requirement to fully represent all substituents and structure permutations in the training set was apparent with the finding of one outlier in the independent test set which was not completely represented in the training set.

Other recent structure-solubility relationship studies have examined even more diverse drug data sets inclusive of compounds containing heterocyclic rings and multiple functional groups [Huuskonen et al., 1998]. Large numbers of topological indices were generated, and sensitivity-based pruning was applied to determine the most influential descriptors. Although the approach and results were sound, applicability to a broader range of chemical structures was restricted because of the limited structure representation in the original data set. By increasing the size of the drug data set, a greater variation in structure was represented. In addition, larger test sets were able to be employed to provide a better estimate of predictive ability of the model [Huuskonen et al., 2000a]. Predictive results were improved but since only 2D electrotopological indices were employed the model was not able to be explained in physical terms.
Once in solution, the penetration of drug molecules across a membrane is the next step in absorption. One study modeled passive drug absorption in rat intestine for a small, structurally diverse series of drugs using immobilised artificial membranes [Genty et al., 2001]. Consistent experimental methods and conditions were ensured by performing in vitro experiments to determine the input descriptors and target absorption values rather than collecting data from the literature. However, the resources required for such experiments were greater than needed for pure in silico modeling. At any rate, improved predictive ability was obtained with the addition of a theoretical descriptor.

Similarly, a combination of experimental and theoretical descriptors has also been used for prediction of human intestinal absorption of drugs [Winiwarter et al., 1998]. The limitation of experimentally determining log $P$ as a descriptor variable meant that only a small training set of compounds was used. The variation in structure was assumed to be representative of a large number of current drugs so prediction of the absorption of independent compounds was reasonable. A larger study using ANNs developed cross-validated models which were then tested with independent compounds [Wessel et al., 1998]. Only theoretical descriptors were utilised in model construction, with absorption data taken from the literature. A number of methods were used to reduce the 162 descriptor set to the final 6, all of which eliminated rather than combined descriptors. Since many of the descriptors were linearly correlated, different final combinations could be found using genetic algorithm selection. Use of cross-validation and independent test sets reduced the training set to 76 compounds for which the model constructed could not be
considered a broadly applicable predictive tool. The model did, however, clearly
differentiate between drugs with high and low absorption values.

A similar study approached prediction of intestinal absorption with a view to
explaining the meaning of theoretical descriptors in the QSPkR model
[Agatonovic-Kustrin et al., 2001]. This was achieved using a radial-basis function
ANN and by generation of descriptors encoding atomistic to 3D holistic
properties. There were 15 descriptors in the optimum model representing a
combination of constitutional, hydrophobic, electronic and steric properties. In
addition to qualitatively indicating absorption characteristics, predictive results
were quantitatively more accurate than in the original study. It was emphasised
that studies based on literature data should be selective to avoid accumulation of
poor or inappropriate data.

A number of structure-bioavailability relationships have been established using
theoretical descriptors. One model constructed for 232 commercial drugs
classified compounds into four classes according to their predicted bioavailability
[Yoshida & Topliss, 2000]. Another QSPkR for 591 compounds developed using
stepwise regression demonstrated that predictions were more accurate than those
achieved using Rule of Five [Andrews et al., 2000]. Both studies included
compounds spanning a broad range of chemical structures which made them
substantially more valuable than models constructed simply from structurally
related compounds. Prediction of bioavailability, as opposed to broad
classification, was performed in the latter and not the former, whereas model
testing using independent compounds was performed in the former and not the
latter. To be of most use in drug development, models should ultimately contain
aspects of both and be developed to quantitatively predict the bioavailability of unknown compounds.

1.6.2 Distribution

Once a drug is absorbed into the systemic circulation its subsequent reversible transfer to extravascular fluids and tissues is termed distribution (Figure 1-5).

Distribution is usually a more rapid process than elimination such that distribution is generally complete while there is still an appreciable amount of drug in the body. Drugs are often bound to proteins in the plasma such as albumin or in tissues. Some drugs distribute preferentially to tissues such as muscle, brain, skin, and fat, or to organs involved in elimination such as the kidney or liver.

![Figure 1-5](image)

**Figure 1-5.** Schematic diagram of a) absorption, b) distribution, and c) elimination in humans (adapted from [Gibaldi, 1984a]).

The partitioning of a drug between tissue and blood is an important pharmacokinetic property describing the distribution of that drug in the body.
under steady-state conditions. Extent of partitioning is given by the partition coefficient, $P_{t,b}$, is defined as follows (Equation 1-2):

$$ P_{t,b} = \frac{C_t}{C_b} $$  \hspace{1cm} \text{Equation 1-2} \\
where $C_t$ is the concentration of drug in the tissue of interest and $C_b$ is the drug concentration in the blood [Shargel & Yu, 1999]. Depending on the characteristics of the drug and availability of plasma pharmacokinetic data, it may sometimes be more appropriate to use the tissue/plasma partition coefficient, $P_{t,p}$.

Several studies have constructed models to predict $P_{t,p}$ from experimental oil/water partitioning and protein binding measurements [Poulin & Theil, 2000; Poulin et al., 2001]. Models were applied to structurally unrelated compounds for a range of tissues in rabbit, rat, mouse and human. The mechanistic nature of the models allowed deductions regarding the effect of lipophilicity on partitioning, as well as causal factors for distribution to particular tissues. Progress was made towards the goal of developing in silico prediction tools from literature data, however the drug data set size limited the general applicability of these models to more diverse chemical entities.

A comparison of mechanistic and ANN methodologies demonstrated that both techniques were able to provide acceptable models for prediction of log $P$ and tissue-to-unbound plasma concentrations for series of analogues [Nestorov et al., 1999]. Physicochemical data was determined experimentally for the construction of both models, however, suitable literature data could have been used instead. Both ANN and PBPK models were constructed using the same descriptive data.
The ANN model provided similarly accurate predictions as the PBPK model did but did not supply any mechanistic information. Should predictions only and not mechanistic information be required, the ANN model could be considered to have equal performance to the PBPK model. Neither of the models was deemed superior and it was suggested that the alternative technique should not be discarded in favour of the other. Instead, they should be used to complement one another.

In another structure-distribution study the distribution of a broad range of drugs into the brain was modeled [Basak et al., 1996]. Input descriptors were topological and E-state indices and the target output parameter was the blood/brain partition coefficient. Variables were manually pruned initially and then subjected to statistical analysis to develop regression equations to predict the blood/brain partition coefficient. In addition to demonstrating their importance in determining tissue distribution, the optimum three-descriptor model allowed the relationship between topological indices and the physicochemical parameters of hydrogen bonding, aromaticity, and molecular branching to be examined.

The apparent volume of distribution does not have a true physiological meaning but represents the theoretical volume into which the drug is distributed. Volume of distribution at steady state can be defined as (Equation 1-3):

\[
V_{ss} = \frac{A_b}{C_p}
\]

\textbf{Equation 1-3}

where \(A_b\) is the amount of drug in the body and \(C_p\) is the concentration of drug in the plasma. High volumes of distribution indicate the preference of a drug to
reside in tissues outside the plasma including erythrocytes and extravascular
tissues. Conversely, low volumes of distribution indicate that a drug is confined
mainly to the plasma [Rowland & Tozer, 1995a]. Apparent volume of
distribution, or more specifically the volume of distribution of the unbound
fraction, can provide useful clinical information since it is generally considered
that the unbound fraction is responsible for the pharmacological action of drug.

Prediction of volume of distribution has been performed using ANNs for 45
structurally unrelated drugs [Ritschel et al., 1995]. Physicochemical parameters of
compounds, allometric data, and theoretical descriptors were used as model
inputs. Validation was performed using a leave-n-out method. Rather than
selecting an optimum set of descriptors, various combinations were evaluated for
predictive ability. It was found that models which included log $P$, protein binding,
and allometric data performed the best, although no detailed analysis of descriptor
significance was performed.

A similar study constructed various QSPkR models for a small series of β-
adrenoceptor antagonists [Gobburu & Shelver, 1995]. Volume of distribution of
the total drug at steady state as well as the volume of distribution of the unbound
fraction at steady state were modeled using ANNs and physicochemical
descriptors taken from the literature. Only one neuron in the hidden layer of the
ANN was required to model the volume of distribution of the unbound fraction.
This indicated a linear relationship between the input and output spaces. Fraction
bound to plasma proteins was also modeled and required a relatively large number
of hidden neurons. Since ANN architecture is generally representative of the
complexity of the parameter being modeled it was surprising to see the fraction
Chapter 1. Introduction

bound requiring five hidden neurons when compared with the single hidden neuron model for volume of distribution of the unbound fraction. Nevertheless, ANN results were an improvement on those obtained using multilinear regression for both training, LOO cross-validation, and testing set predictions.

1.6.3 Metabolism and Excretion

Elimination of drugs for the body occurs via the processes of metabolism and excretion (Figure 1-6).

![Figure 1-6. Drug elimination depicted schematically, showing processes of a) metabolism and b) excretion (adapted from [Rowland & Tozer, 1995b]).](image)

Drug metabolism, or biotransformation, is the chemical conversion of the parent species into one of a number of metabolites. Drugs are excreted from the body as either the parent compound or as one of the metabolites. Excretion is primarily via the renal pathway via the urine, and also through the hepatobiliary route via the faeces. For volatile compounds it may be through the breath, and occasionally drugs are also excreted in the sweat. These are mainly Phase I reduction, oxidation, or hydrolysis reactions and/or Phase II conjugation reactions. Enzymes involved in metabolism are located predominantly in the liver, although other sites
Chapter 1. Introduction

of enzymatic metabolism include the intestinal wall, kidney, lung, and skin. The major class involved in metabolism is the cytochrome P450 (CYP) enzyme superfamily.

CYP isozymes often display structural specificity in metabolic reactions. Approaching metabolism from a modeling perspective is greatly challenging due to the complexity of the metabolic system in its entirety. Different drug classes can be metabolised by different enzyme classes, and the same is often true even for structurally related compounds. Moreover, it is not uncommon for a drug to be metabolised by simultaneous and competing enzymatic pathways. The extent of such metabolism depends on the individual rate constants of the competing pathways. To further complicate matters, expression and morphology of metabolic enzymes can show large variations across ethnic groups.

One approach in structure-metabolism relationship modeling has been to characterise molecules or structural motifs likely to interact with one specific enzyme [Mlinsek et al., 2001]. In doing so, the problem is simplified to dealing with a known target structure composed of one or more identifiable active sites. Thus, characterisation of the structural characteristics of theoretical ligands which affect binding is more easily performed. A study examining flavonoid derivatives made use of quantum chemical descriptors to model the inhibition of CYP 1A2 [Moon et al., 2000]. 3D conformation of ligands is crucial for enzyme binding so all structures were presented as energy-minimised conformers. Both multilinear regression and ANN models were constructed, with one ANN model making use of the same descriptors as the multilinear regression model. Direct comparison of predictive ability of each technique was thus possible, and since nonlinear
relationships were assumed in flavonoid-CYP 1A2 binding, the ANN provided superior results. Another subset of descriptors was determined for the ANN using sensitivity-based pruning. This subset allowed more accurate predictions for the test set of compounds. This finding demonstrated the need for nonlinear methods of descriptor selection to be employed for nonlinear systems.

A more difficult approach is to model the enzymatic biotransformation of drugs by possibly numerous potential enzyme species. Complexity can be reduced by examination of just a single enzyme class instead of looking broadly at multiple enzyme classes. Carboxylic ester hydrolases catalyse the hydrolysis of a variety of ester-containing substances and are present in many human tissues. They have broad substrate specificity towards esters and amides and it is known that a single drug compound can be hydrolysed by more than one particular enzyme. The in \textit{vitro} metabolism of a number of structurally unrelated compounds was modeled based on structural characteristics [Buchwald & Bodor, 1999]. Conformation of substrate drugs was determined by rigorous \textit{in silico} minimisation routines. From the minimised structures quantum chemical numbers, calculated log \textit{P} descriptors, and a novel descriptor representing a theoretical steric angle were generated. Even though training results were acceptable the prediction of enzymatic metabolism of independent test compounds proved more difficult. The final models qualitatively differentiated between slowly and quickly metabolised substrates but quantitative predictions varied considerably from experimental values. An accurate predictive model was not expected, however, due to the size of the training data set and complexity of enzymatic metabolism.
In a different problem, classification of metabolic fate of drugs has been performed for structurally related compounds. One study examined urinary excretion of glucuronide conjugates, glycine conjugates, and unchanged parent drug for 22 benzoic acid derivatives [Cupid et al., 1999]. Information regarding the metabolic enzymes was neither required nor explored since the study was not mechanistic by nature. Descriptors were calculated from structure and included quantum chemical numbers, geometrical descriptors, and partition coefficients. Predictive capability determined using a leave-2-out procedure indicated that structure-metabolism relationships were able to be modeled for the structurally-related compounds examined.

1.6.3.1 Clearance

Clearance is defined as the volume of blood cleared of drug per unit time. It is a function of both the intrinsic ability of eliminating organs such as the liver and kidney to excrete or metabolise a drug, and the blood flow rate to these organs (Figure 1-7). Clearance due to a single organ is given as the product of the blood flow to that organ and the extraction ratio (Equation 1-4):

\[ CL = Q \cdot \frac{C_A - C_V}{C_A} = Q \cdot ER \]  

where \( Q \) is the blood flow to the organ, \( C_A \) is the concentration of drug in the arterial blood, \( C_V \) is the concentration of drug in the venous blood, and \( ER \) is the extraction ratio. The extraction ratio is the ratio of the rate of elimination of a drug to the input rate of the drug to an organ. Thus, the higher the extraction ratio the more drug is eliminated and the less passes through the eliminating organ intact.
Total clearance is a complex parameter since it combines elements of both metabolism and excretion. A simpler parameter is renal clearance which mainly involves the processes of glomerular filtration and reabsorption. A comparison of models for renal and nonrenal clearance for a series of $\beta$-adrenoceptor antagonists gave similar cross-validation results for both [Gobburu & Shelver, 1995]. Even though this was not the expected result it was most likely due to the small size of the data set. When clearances of independent compounds were examined, only the prediction of renal clearance was acceptable whereas prediction of nonrenal clearance was not.

**Figure 1-7.** Schematic diagram of drug elimination by a single organ (adapted from [Gibaldi, 1984b]).

In a related study a slightly larger data set of more structurally diverse drugs was used to develop a model for *in vivo* hepatic clearance [Schneider et al., 1999]. The aim was to compare different modeling techniques to determine the feasibility of developing structure-clearance relationship models. A combination of allometric and *in vitro* data were used as input variables. It was found that adequate cross-validated models could be achieved using *in vitro* data only. Even though the data set was relatively small a variety of metabolic pathways were represented covering both Phase I and Phase II reactions. Nevertheless, more broadly
applicable models would require much larger data sets to cover the large range of metabolic pathways and also to allow true predictive performance using independent compounds to be assessed.

A structure-clearance relationship was developed in conjunction with the structure-distribution relationship mentioned in Section 1.6.2 [Ritschel et al., 1995]. The ANN model was novel in that two pharmacokinetic parameters, clearance and volume of distribution, were predicted simultaneously. Models were cross-validated but testing of independent drugs was not performed. Models were for the most part rather inaccurate. Moreover, quantitative structure-pharmacokinetic relationships were not examined. It was likely that the information content of the descriptors was inadequate to predict two complex pharmacokinetic parameters at once.

1.7 Summary Remarks

A review of the literature revealed that predictive QSPkR modeling is a relatively undeveloped area in which there is ample scope for progress using newer soft computing techniques such as ANNs. A number of methodologies have been applied from both mechanistic and non-mechanistic approaches. Should solely predictive models be required then a non-mechanistic approach would seem appropriate. To aid in drug development, however, information should be taken from the QSPkR models regarding the quantitative effects of structure on NCE pharmacokinetics. QSPkRs that have been developed to date have mainly used physicochemical or experimentally-derived parameters to construct models. Theoretical descriptors offer a quicker and more effective alternative. Each
theoretical descriptor provides a certain amount of unique information. Modeling techniques need to be used that allow selection of the most appropriate set of descriptors for optimum predictive capabilities. Moreover, a combination of descriptors should be used initially to encode as much of the multidimensional nature of a chemical structure as possible.

ANNs have thus far only been used to a small extent in developing QSPkRs. They are a robust modeling tool and have certain speed and nonlinearity advantages where other methods do not. Construction of ANN-based QSPkRs should be demonstrated to be effective in simple systems first such as for structurally related sets of compounds, and then extended to include larger numbers of structurally diverse compounds. All models should be validated with an appropriate technique to ensure adequate predictive capability. In the literature to date most studies have addressed some of these issues but not all have to any great extent. The relatively small number of studies and incomplete nature of those done has provided the opportunity for the present research to be undertaken. The work presented in this dissertation validates the applicability of the present technique for both small and large sets of data, and in doing so addresses the issue of congeneric and structurally diverse drug data sets. The modeling described takes its basis from conventional QSPkR modeling published using both physicochemical and theoretical descriptors. All models were appropriately validated to ensure reliability in contrast to some earlier studies. In utilising ANNs and purely theoretical descriptors this thesis has advanced the field of QSPkR modeling with a view to aiding the drug development process in the early stages.
Chapter 2

General Methodology
Chapter 2. General Methodology

2.1 Pharmacokinetic Data ................................................................. 70
2.2 Molecular Structure Data ............................................................ 70
2.3 Descriptor Generation ............................................................... 71
2.4 Descriptor Selection ................................................................. 72
2.5 Model Construction ................................................................. 73
  2.5.1 ANN Training ................................................................. 73
  2.5.1.1 Multilayer perceptron ................................................ 73
  2.5.1.2 Radial-basis function ................................................ 74
  2.5.2 Model Validation .............................................................. 75
2.6 Operating Characteristics of ANNs .............................................. 76
  2.6.1 Multilayer Perceptron ....................................................... 76
    2.6.1.1 Neurons ................................................................. 77
    2.6.1.2 Learning Rule ......................................................... 78
  2.6.2 Radial-Basis Function ANNs ............................................ 80
    2.6.2.1 Neurons ................................................................. 80
    2.6.2.2 Kernel Function ...................................................... 80
    2.6.2.3 Model Optimisation ................................................ 82
Chapter 2. General Methodology

A flow-chart describing the general aspects of QSPkR modeling is given below (Figure 2-8). Important issues involved in the work presented in this thesis are discussed in the following sections.

![Flowchart of QSPkR modeling process]

**Figure 2-8.** General outline of the QSPkR modeling process.
2.1 Pharmacokinetic Data

Pharmacokinetic data for all studies were taken from the literature. The congeneric adenosine $A_1$ receptor agonist and $\beta$-adrenoceptor antagonist data for Chapter 3 and Chapter 4 respectively were taken directly from studies published using the same data. This enabled comparisons to be made with the original studies while at the same time offering a degree of consistency amongst the data. Data for cephalosporins (Chapter 5) as well as pharmacokinetic data for the structurally diverse drugs used in Chapter 6 were collected from the literature after careful screening. Only actual values, and not relative values, of bioavailability were accepted for Chapter 6. Pharmacokinetic values for Chapter 5 were converted into the units given in Table 2-4. A bodyweight of 70 kg was assumed where bodyweights were not given in the literature.

Table 2-4. Units for pharmacokinetic parameters examined in Chapter 5.

<table>
<thead>
<tr>
<th>Pharmacokinetic parameter</th>
<th>Symbol</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clearance (renal)</td>
<td>$CL_R$</td>
<td>mL·min$^{-1}$·kg$^{-1}$</td>
</tr>
<tr>
<td>Clearance (total)</td>
<td>$CL$</td>
<td>mL·min$^{-1}$·kg$^{-1}$</td>
</tr>
<tr>
<td>Fraction bound to plasma proteins</td>
<td>$f_b$</td>
<td>N/A</td>
</tr>
<tr>
<td>Fraction excreted in urine</td>
<td>$f_e$</td>
<td>N/A</td>
</tr>
<tr>
<td>Half life</td>
<td>$t_{1/2}$</td>
<td>h</td>
</tr>
<tr>
<td>Volume of distribution at steady state</td>
<td>$V_{ss}$</td>
<td>L·kg$^{-1}$</td>
</tr>
</tbody>
</table>

2.2 Molecular Structure Data

Structures for all compounds were taken from the Merck Index [Budavari, 1996] and modeled in Accumodel 1.0 [MicroSimulations, 1996] in a Windows
Chapter 2. General Methodology

[Microsoft, 1998b] environment. Where only constitutional or 2D descriptors were required, the MM3 energy minimisation routine in Accumodel was applied. Output files were saved in native MMF format. In Chapter 6, 3D descriptors were used and accurate energy-minimised structures were required. Structures from Accumodel were exported to MOL format and then imported into Catalyst 3.1 [MSI Molecular Simulations, 1996] on a Unix machine. Structures were re-verified and nominally energy minimised using the 3D-minimise routine in Catalyst. Structures were then exported to MOL2 format and imported into Sybyl 6.7 [Tripos, 2000]. In Sybyl, structures were energy minimised according to the routine described in Appendix A1 (Appendices provided on enclosed CD). Following optimisation of 3D conformation, structures were saved and imported into Catalyst and then re-exported to SDF format. SDF format allowed multiple compounds to exist in the one file.

2.3 Descriptor Generation

Programming code for calculation of constitutional descriptors was written in Microsoft Visual Basic 6.0 [Microsoft, 1998a] for use in Windows. MMF files were used as input data since these descriptors were conformation-independent. Output files were ASCII format and additional manipulation was required in Microsoft Excel to extract the data. Programming code for generation of connectivity indices and their linear combinations, charge indices, and other topological indices used in Chapter 3 and 4 was written and kindly supplied by Mostafa Hosseini.
Descriptors dependent on 3D conformation used in Chapters 5 and 6 were generated from the optimised structures as determined in Sybyl. Quantum chemical numbers were generated in Sybyl using the normal routine constrained to precise calculation. Full optimisation was included to account for subtle effects of conformation. An additional keyword, NOMM, was included to permit calculations to be made for compounds containing nitro groups. Full details of minimisation routines are given in Appendix A1. Other software, Molecular Modeling Pro Demo 4.07 [ChemSW, 2001] and CAChe Project leader 3.11 [Oxford Molecular, 2001] were used to calculate descriptors used in Chapter 5 and Chapter 6 from MOL files.

2.4 Descriptor Selection

Manual selective pruning involves examination of ANN weight matrices as well as cross-correlation of descriptors. Pruning is performed by first removing highly correlated input variables and then removing those assigned low absolute weights. This method eliminates redundant information carried in correlated descriptors, a factor which may be overlooked by some sensitivity-based methods. In addition, selective pruning accounts for descriptors with low weight which may still contribute significantly to the model and which would otherwise be eliminated by penalty term algorithms.

In a similar manner to genetic algorithms, sensitivity-based pruning allows for faster, automated pruning to take place. Sensitivity ratios represent the reliance of a model on the information contained in each input variable. The sensitivity value for each descriptor was calculated in the following manner: information contained
in the descriptor was substituted with meaningless values and the network was retrained. The error associated with the retrained network was then compared with the baseline error of the optimum model. Sensitivity of the descriptor was defined as the ratio of the former to the latter. Hence, sensitivities greater than one indicated that the descriptor provided useful information and removal of that descriptor would be detrimental to the model. Higher sensitivities correspond to a greater reliance of the model on the information content of the corresponding descriptor. Based on results of sensitivity analyses, inputs with sensitivities less than one are usually eliminated from the model. It is important to note that sensitivity ratios do not define the absolute importance of a descriptor. Descriptors are often interdependent upon one another and subtle combinations may be required to properly represent a given situation. Hence, whilst sensitivity values can indicate the importance of certain descriptors they do not determine the absolute usefulness or redundancy of any single descriptor in a model.

2.5 Model Construction

2.5.1 ANN Training

2.5.1.1 Multilayer perceptron

The multilayer perceptron, a feed-forward back-propagation ANN, has a customisable architecture allowing zero, one, or multiple hidden layers of neurons. Prior to prediction these ANNs must undergo supervised training.

The values of the weights in untrained ANNs are initially random, and then are modified during the course of training [Zupan & Gasteiger, 1999]. The process of
supervised training involves presentation of both input and output data to the ANN. The input data is fed into the network which produces a prediction as its final output (feed-forward). The network then calculates the error between its predicted output and the actual output, and proceeds to modify the weights of the network in order to reduce this calculated error (back-propagation). This process is repeated until the weights of the network are optimised such that the error between the actual and predicted outputs is minimised to a predetermined value. Each repetition is termed a cycle, and thousands of these cycles may be required to complete training. The most common method for back-propagation in the multilayer perceptron is the generalised delta rule, of which a detailed description is given in Section 2.6.1.

2.5.1.2 Radial-basis function

During training of a radial-basis function ANN the response surface is modeled by the Gaussian transfer functions in the hidden layer. The shape of the response surface can be modified for each hidden neuron to eventually provide a generalised representation of the output space. Optimisable parameters include height, slope and “flatness” (Figure 2-9), as well as the position of the centre of the curve [Lohninger, 1993]. Network architecture is limited to only three layers of neurons, although additional bias and indicator neurons may also be included. Functionally, radial-basis function ANNs operate in a similar manner to multilayer perceptrons in that training data is fed in a forwards direction through the ANN after which a prediction is made. Following this, the weights in the radial-basis function ANN are adjusted to provide a solution for modeling of the output space by the hidden layer Gaussian functions. A more detailed description
of the operational characteristics of radial-basis function ANNs is given in Section 2.6.2.

**Figure 2-9.** Optimisable parameters of the Gaussian kernel function in radial-basis function ANNs: a) height, b) slope, and c) flatness.

### 2.5.2 Model Validation

In small data sets leave-one-out (LOO) cross-validation is normally applied. Under this system, a model is trained with all compounds except for one which is left out. The trained model is then used to predict the target output of the compound that was excluded. The process is repeated with the next compound
being left out instead, and so on until each compound has been left out once. All predictions are then compared with observed values to give an indication of predictive power of the model.

When more data is available such as in medium to large drug data sets, compounds can be divided into a training set, validation set and testing set. Training is performed using the training set, and model performance is monitored with the validation set. Ultimate predictive ability is then determined by prediction of the target outputs of the independent testing set.

2.6 Operating Characteristics of ANNs

The following section provides a more detailed description of the algorithms and processes utilised in ANN training. Only the methods used for the present research will be described.

2.6.1 Multilayer Perceptron

These ANNs may contain multiple hidden layers of neurons with each neuron having the potential to perform a nonlinear transformation of data. Even so, the input layer in a multilayer perceptron does not usually nonlinearily transform the data. Instead, it linearly scales the data and serves a distributive role by introducing the values of the input variables into the network for subsequent nonlinear transformation in the hidden and output layers.
2.6.1.1 Neurons

There are four operations performed by each neuron in the hidden and output layers. These are the input/output function, summation function, activation function, and the transfer function. The input/output function evaluates the input signals to the neuron from the neurons in the previous layer and also passes the output signal to the neurons in the following layer. The actual input received is equal to the output from the previous neuron multiplied by the weight of the connection. All the inputs from the previous layer are then added together by the summation function. Summation proceeds according to (Equation 2-5):

\[ I_{\lambda,i} = \sum_j w_{ij} O_{\lambda-1,j} + B_{\lambda} \]  

Equation 2-5

where \( I \) is the total input, \( w \) is the weight, \( O \) is the output, \( i \) and \( j \) are neurons in the \( \lambda \) and \( \lambda-1 \) layers respectively, and \( B \) is the bias to each layer which has no inputs from the previous layer (Figure 1-2). This sum is then passed on to the activation function which allows the sum to vary with respect to time before it is given on to the transfer function. The transfer function “squashes” the result of summation onto a continuous nonlinear curve between the values zero and one, for example a sigmoid curve (Figure 1-3b).

This result is passed on to the output path which may be either to the next layer of neurons, or the output path to the final result or prediction. Most commonly the transfer function is sigmoidal (Equation 2-6) and for multilayer perceptron ANNs the output of each neuron must be a differentiable and continuous signal for the learning process to occur.
Between the neurons $i$ and $j$ there exists an hypothetical connection formed by the relative weight value. This weight determines the extent to which the output from $j$ influences the final output of the network.

$$O_{i,j} = f(I_{i,j}) = \frac{1}{1 + \exp^{-I_{i,j}}}$$  \hspace{1cm} \text{Equation 2-6}

Multiplication of the output by the weight produces a scalar value which provides the following neuron with information as to the importance of that output. Since weights can take either positive or negative values then, analogous to biological systems, negative weights may be inhibitory and positive weights excitatory. In multilayer perceptrons with a predetermined number of layers and neurons the weights are the only parameters adjusted. This is the basis for the assertion made earlier that the total knowledge contained in the network is stored in the strengths of the connections between the neurons and not in the individual neurons themselves.

2.6.1.2 \hspace{0.5cm} \textit{Learning Rule}

Error reduction via a back-propagation algorithm will now be described. After the feed-forward process the error, $\delta$, occurring in the output layer is calculated according to (Equation 2-7):

$$\delta_{i,j} = (O'_{i,j} - O^*_{i,j})f'(I_{i,j})$$  \hspace{1cm} \text{Equation 2-7}

where $O'$ is the target output and $O^*$ is the actual output. Now the need for a continuous and differentiable transfer function is evident since $f'(I_{i,j})$ is the first derivative of this function. The target value is unknown for the neurons in the
hidden layer so it is calculated with respect to the errors in the output layer proceeding it (Equation 2-8):

\[ \delta_{i,j} = \sum_j \left( \delta_{i+1,j} w_{j} \right) f'(I_{i,j}) \]  

Equation 2-8

Now the weight between the neurons \( i \) and \( j \) in the hidden and output layers respectively is adjusted such that the magnitude of the adjustment is given by (Equation 2-9):

\[ \Delta w_{ji} = \eta \delta_{ji} O_{i-1,j} + \mu \Delta_{n-1} w_{ji} \]  

Equation 2-9

where \( \eta \) is the adjustable learning rate, \( \mu \) is a momentum term, and \( n \) is the number of cycles run by the network. In order to achieve smooth reduction of error the learning rate must be extremely small. This, however, gives rise to the problem of local minima: the network may get stuck in a local error minimum and so not reach the global error minimum. It is for this reason that the momentum term is added since it helps the network carry the weights past local minima. Equation 2-9 is known as the generalised delta rule [Aston & Wilding, 1992].

In summary, training is the minimisation of error using the training data and is achieved by adjustments of the weight values. Training of multilayer perceptrons is performed layer-by-layer in a back-propagating fashion according to the generalised delta rule.
2.6.2 **Radial-Basis Function ANNs**

Radial-basis function ANNs are capable of mimicking a multilayer perceptron models, and the opposite is true also [Haykin, 1994]. They are, however, operationally distinct from one another.

### 2.6.2.1 Neurons

Radial-basis function ANNs differ from multilayer perceptrons in that the nonlinear transformation of data is performed only in the hidden layer of neurons. The input layer of neurons still serves a scaling and distributive role for the descriptor data. The hidden layer contains neurons with a Gaussian kernel function as the transfer function (Figure 1-3a). The output layer simply summates the resultant transformed values from the hidden layer. The power of radial-basis function ANNs lies in the multiple adjustable Gaussian functions given in the hidden layer which model the multidimensional output space (Figure 2-9).

### 2.6.2.2 Kernel Function

The generalised Gaussian kernel function in neuron $j$ is given by (Equation 2-10):

$$
\varphi(x_j) = \frac{1}{\exp\left(\sigma(x_j - c_j)^T A(x_j - c_j)\right)}
$$

**Equation 2-10**

where $c$ is the centre position of the kernel function (centroid) in the $n$-dimensional space and $\sigma$ is a scaling factor [Lohninger, 1993]. The vector $x$ holds the input data to neuron $j$. The matrix $A$ serves to normalise the input data space and, if $A$ is an identity matrix, can be substituted with the Euclidean distances between the input and centroid ($x_j$ and $c_j$ respectively). The superscript $T$ simply
denotes transposition of the vectors. One adjustable parameter is the centroid value \( c_j \) which determines the geometric centre of the curve since Gaussian functions are symmetrical. Modification of \( \sigma \) changes the shape of the kernel function which thus facilitates fitting of the curve to the output space.

Further utility from the kernel function can be gained by including a flatness parameter \( \phi \) (Equation 2-11):

\[
\varphi(x_j) = \frac{1 + \phi}{\phi + \exp\left[\frac{\sigma(x_j - c_j)^T A(x_j - c_j)}{\phi}\right]}
\]

Equation 2-11

This allows flattening of the kernel function at the top and also broadening the width of the curve.

The Gaussian function thus obtained (Equation 2-11) forms a basis for the optimisation of the approximation function. Solving the approximation function involves modification of the adjustable parameters to best model the output space.

The approximation function is given by (Equation 2-12):

\[
f(x) = \sum_{j=1}^{h} w_j \varphi(x_j - c_j)
\]

Equation 2-12

where \( h \) is the number of hidden neurons and hence the number of kernel functions, and \( w \) represents the weights to the output layer.

The number of hidden neurons should be less than the number of training patterns in order for generalisation to take place. If the numbers of hidden neurons and
training patterns are equal then Equation 2-12 can be solved exactly and the data will be overfitted.

2.6.2.3 Model Optimisation

Optimisable parameters in Equation 2-12 include $h$, $c$, $\sigma$, $\phi$, and $w$. Optimisation is initially performed with respect to the characteristics of the data set.

The number of hidden neurons, $h$, is determined first. In a similar manner to multilayer perceptrons the smallest number of neurons able to provide adequate complexity for the model is used.

Next the centroids, $c$, of the hidden neurons are established. A common method is by randomly sub-sampling the training set and mapping the hidden neurons to those positions. Since $h$ is less than the number of patterns and the target space is sampled randomly, then the distribution of the centroids should provide a general representation of the output space. A more accurate method of seeding the centroids is by the K-means algorithm. This involves clustering of the training data and then positioning of the centroids to most suitably represent the clustered data points. The criteria for the K-means clustering are that each data point belongs to a cluster centre and is nearer to this centre than to any other centre, and that each cluster centre is the centroid of the data points which belong to it.

Once the centroids have been established the curve shape characteristics, $\sigma$ and $\phi$, can be set. In order to model finer details the kernel function can be made narrow and sharp, however this can lead to overfitting. In order to generalise the kernel function can be made flat and wide, which can instead lead to underfitting. Initially a compromise is struck between the two, with kernel functions typically...
overlapping neighbouring functions. A simple way of setting the shape, known as the isotropic method, is to assign all hidden neurons identical curves. The shapes are chosen with respect to the number of centres and the volume of space they occupy. A more rigorous method of assigning initial shape characteristics is the K-nearest neighbour technique. This approach examines the clustering of the hidden neurons and target space. The size and shape of each kernel function is set to reflect the density of data around it. Thus, in areas of high data density the kernel functions are smaller and sharper to preserve detail. Conversely, in areas of sparser data the kernel functions are more expansive and broad. The K-nearest neighbour technique is more computationally expensive than the isotropic method.

Once the kernel functions have been seeded and defined then training can proceed. The system can be reduced to the following form (Equation 2-13):

\[ y = Dw \]  

\textbf{Equation 2-13}

where \( y \) is the vector of the target values of all training patterns, \( w \) is the vector of the weights \( w_j \), and \( D \) is a design matrix containing the elements given in Equation 2-11. A column vector with all elements set to one can also be added to \( D \) to incorporate a bias neuron in the output layer.

Equation 2-13 can then be solved for \( w \) by using the pseudo-inverse with singular value decomposition algorithm (Equation 2-14):

\[ w = (D^T D)^{-1} D^T y \]  

\textbf{Equation 2-14}
In summary, training of a radial-basis function ANN involves seeding of appropriate numbers of kernel functions to adequately reflect the training space. Parameters governing the kernel function can be modified. The optimum model is determined by solving the value of the weights in the ANN for the given kernel functions and target output values.
Chapter 3. Selective Descriptor Pruning for QSPR Studies Using ANNs
Chapter 3. Selective Descriptor Pruning for QSPR Studies Using ANNs

3.1 INTRODUCTION ..................................................................................................................87
3.1.1 Descriptor Selection ...........................................................................................................87
3.1.2 Patterns to Connections Ratio – $\rho$ ..............................................................................88
3.1.3 Study Aims .........................................................................................................................89

3.2 METHODS ..............................................................................................................................89
3.2.1 Linear Artificial Structured Data.......................................................................................90
   3.2.1.1 Simulation of Experimental Error ..............................................................................91
3.2.2 Nonlinear Artificial Structured Data ...............................................................................91
3.2.3 Time-Dependent Artificial Structured Data .................................................................91
3.2.4 Testing Data ......................................................................................................................92
3.2.5 Literature Experimental Data .........................................................................................93
3.2.6 Artificial Neural Network Model ....................................................................................94
   3.2.6.1 Model Parameters .......................................................................................................94
   3.2.6.2 Model Training ..........................................................................................................95
   3.2.6.3 Pruning of Descriptors ..............................................................................................95
3.2.7 Signal:noise and $\rho$ .......................................................................................................96

3.3 RESULTS AND DISCUSSION ................................................................................................97
3.3.1 Relevance of Data .............................................................................................................97
3.3.2 Signal:Noise Ratio ...........................................................................................................98
   3.3.2.1 Linear and Time-Dependent Data ..............................................................................99
   3.3.2.2 Nonlinear Data ........................................................................................................100
   3.3.2.3 Literature Experimental Data ..................................................................................102
3.3.3 Model Predictions and $\rho$ .............................................................................................104
   3.3.3.1 Linear Predictions ......................................................................................................104
   3.3.3.2 Nonlinear Predictions ...............................................................................................105
   3.3.3.3 Time-Dependent Predictions ..................................................................................106
   3.3.3.4 Predictions for Experimental Data ...........................................................................108
3.3.4 The Effect of $\rho$ .............................................................................................................110

3.4 CONCLUSIONS .....................................................................................................................111
Chapter 3. Selective Descriptor Pruning for QSPR Studies Using ANNs

3.1 Introduction

ANNs have been used in drug development mainly for structure-activity relationship (QSAR) analyses. They have also been used for the construction of structure-property relationship (QSPR) models, under which structure-pharmacokinetic modeling falls. In this chapter the operating characteristics of ANNs are investigated with different representative types of data. This is done to justify the validity of selective descriptor pruning for ANN models in pharmacokinetics. The results presented can be broadly applied to both QSAR and QSPR studies using ANNs. Work described in this chapter has been accepted for publication in the Journal of Computational Chemistry.

3.1.1 Descriptor Selection

Selection of optimal descriptors in QSAR and QSPR studies has been a perennial problem. Various selection or pruning techniques exist which aim to eliminate redundant weights and/or descriptors leaving only those offering a significant contribution to the model. Both manual and sensitivity based pruning have been discussed in Section 1.4.4 and Section 2.4. Manual selective pruning of single and multiple descriptors has been used successfully to reduce the number of model input variables [Maddalena & Johnston, 1995]. Conveniently, manual selective pruning allows examination of weight matrices at each step of the process.
3.1.2 Patterns to Connections Ratio – Rho

In any ANN model the sample size (number of patterns) can be related to number of connections by the parameter $\rho$, which is defined as the ratio of the former to the latter (Equation 3-15):

$$\rho = \frac{P}{(I+1) \times H + (H+1) \times O}$$

Equation 3-15

where $P$ is the number of patterns, $I$ is the number of input variables, $H$ is the number of hidden neurons, and $O$ is the number of output neurons. Equation 3-15 also allows for the bias connection to the input and hidden layers [Kovesdi et al., 1999]. Several studies [Andrea & Kalayeh, 1991; So & Richards, 1992] have examined $\rho$ in order to provide guidelines for optimum network architecture. Following work on a series of dihydrofolate reductase inhibitors Andrea & Kalayeh (1991) claimed that optimum ANN models had the range $1.8 < \rho < 2.2$. For $\rho > 2.2$ there were insufficient connections for the ANN to extract necessary information hence predictive performance was poor. ANN predictions also suffered for $\rho < 1.8$ because training data was overfitted. They claimed that as a general rule for ANN studies $\rho$ should be greater than one. So & Richards (1992) observed that for $\rho >> 3.0$ ANNs were unable to generalise and that the actual effective range of $\rho$ may be implementation-dependent.

In an attempt to examine the effect of $\rho$ more closely one study took both artificial structured (AS) data sets as well as examples of real QSAR data from the literature and varied the number of hidden neurons in ANN models [Manallack et al., 1994]. Results indicated that at low values of $\rho$ ANN training correlations
were high but predictive ability as tested by leave-one-out (LOO) cross-validation (see Section 2.5.2) was poor. Network architecture was critical to ANN performance but general guidelines for experimental data sets could not be established. There have been few studies investigating the relationship between \( \rho \) and predictive performance by varying sample size with a fixed number of hidden neurons.

### 3.1.3 Study Aims

The purpose of the current study was twofold. First, selective descriptor pruning as it applies to QSPkRs was examined. This was done using large populations of real and synthetic input descriptors. Second, the effects of pruning on the parameter \( \rho \) was investigated to assess any relationship with predictive performance.

### 3.2 Methods

Two types of data were used. AS data sets were constructed for each of linear, nonlinear, and time-dependent data formats. AS data sets were designed to be representative of the different types of data encountered in structure-activity and structure-property relationship studies. The second type of data used was real, experimental data taken from the literature. Trends and relationships were examined with the AS data and verified with the literature data.
Chapter 3. Selective Descriptor Pruning

3.2.1 Linear Artificial Structured Data

The linear AS data sets consisted of matrices of data designed to represent a linear situation where only one descriptor variable was correlated with the target property (Figure 3-10). Each matrix consisted of a column of patterns which represented individual drugs. Next was an independent “meaningful” column (X) and a dependent column (Y), where Y was a linear function of X. Thus the Y data represented output property values for each of the patterns.

<table>
<thead>
<tr>
<th>Pattern</th>
<th>X (meaningful input variable)</th>
<th>R₁</th>
<th>R₂</th>
<th>.</th>
<th>R₉₉</th>
<th>Y (dependent output)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>2</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>3</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>20</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
</tbody>
</table>

**Figure 3-10.** Data matrix for linear AS data with 100 input variables and 20 patterns. Patterns were arranged in rows and variables were arranged in columns. Input variables consisted of one meaningful variable (X) and 99 meaningless variables (R₁ to R₉₉). There was one dependent output variable (Y).

The “meaningless” data (R₁ – R₉₉) was constructed such that each value generated was a uniformly distributed random number between +1.0 and -0.9. Data sets contained up to a maximum of 100 columns of these random input variables representing useless descriptors.

Linear AS data sets contained 20 patterns with the values for the dependent Y variable being equal to the corresponding values for the meaningful X variable.
3.2.1.1 Simulation of Experimental Error

There is a certain amount of error associated with values obtained by experimental means. In order to simulate such error another linear AS data set was generated which contained a quantity of “noise.” A random number between +0.2 and -0.2 was generated and added to each of the X values. Since the meaningful X values in the pure linear AS data set had a maximum absolute value of one, then the new values for the X variable represented experimental data with up to 20% variation.

3.2.2 Nonlinear Artificial Structured Data

A simple way to incorporate nonlinearity in multilinear regression equations is to include a squared-term of one of the descriptor variables. ANNs have no such need to make provision for nonlinearity. They are useful for managing nonlinear relationships amongst data, and this was examined by defining a parabolic relationship between the dependent target variable and independent meaningful variable in an AS data set.

Generation of the data matrix followed a similar procedure to that described in Section 3.2.1. The major difference was that values for Y were equal to the square of corresponding X values. In addition, there were 40 patterns in the nonlinear AS data set instead of 20, and upper and lower limits of the X variable were +1.00 and -0.95 respectively. A noisy nonlinear AS data set containing up to 20% variation was also generated.

3.2.3 Time-Dependent Artificial Structured Data

To model time-dependency the output Y variable was represented by a periodic function. A sine curve was chosen and values for the target Y variable described a
curve $5.58\pi$ radians in length. An AS data matrix was then constructed in which there were two meaningful variables, $X_1$ and $X_2$ (Figure 3-11). Values for the meaningful variables were calculated by doing a phase shift of $Y$. Values of $X_1$ were equal to the corresponding $Y$ value plus $0.4\pi$ radians, and values of $X_2$ were equal to the corresponding $Y$ value plus $0.5\pi$ radians. Both variables $X_1$ and $X_2$ were examined for correlation and neither was found to be linearly dependent on each other or on the output variable $Y$.

<table>
<thead>
<tr>
<th>Pattern</th>
<th>$X_1$</th>
<th>$X_2$</th>
<th>$R_1$</th>
<th>$R_2$</th>
<th>$\ldots$</th>
<th>$R_{78}$</th>
<th>$Y$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>2</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>3</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>55</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
</tbody>
</table>

Figure 3-11. Data matrix for time-dependent AS data with 80 input variables and 55 patterns. Patterns were arranged in rows and variables were arranged in columns. There were two meaningful descriptors ($X_1$ and $X_2$) and 78 meaningless input variables ($R_1$ to $R_{78}$) for the one target output ($Y$).

### 3.2.4 Testing Data

The format of the testing data was the same as for the training data for the linear, nonlinear, and time-dependent AS data sets, except that there was no target output variable. The linear and noisy linear testing sets were the same as each other. Linear testing data consisted of a meaningful variable, $X$, and the same number of meaningless variables as the corresponding training set. Values for $X$ consisted of 20 evenly distributed numbers lying between $0.95$ and $-0.95$. For this range all
the testing data except for one point lay within the bounds of the training data to allow interpolation by the ANN model. The proposed target output variable, \( Y \), was equal to the meaningful variable, \( X \).

Nonlinear test data was generated in a similar manner as linear data. There were 40 patterns with an upper and lower limit of 0.98 and -0.97 respectively for the meaningful \( X \) variable. The proposed target output space, \( Y \), was the square of the corresponding \( X \) values.

Time-dependent test data was constructed with two meaningful variables and the same number of random variables as the corresponding training AS data set. In the testing set, values of \( X_1 \) were equal to the corresponding \( Y \) value plus 0.1\( \pi \) radians, and values of \( X_2 \) were equal to the corresponding \( Y \) value plus 0.2\( \pi \) radians. Neither \( X_1 \) nor \( X_2 \) in the testing set were linearly correlated with each other or the proposed target output space.

### 3.2.5 Literature Experimental Data

A set of 1,4-benzodiazepin-2-ones taken from the literature [Maddalena & Johnston, 1995] was used to train the ANN during the pruning and cross-validation study. This data set consisted of 57 individual 1,4-benzodiazepin-2-ones substituted at positions R7, R1, R2’, R6’, R3 and R8, and their associated log \( IC_{50} \) values. Each of the six positions was described by lipophilicity \( (\pi) \), molar refractivity \( (MR) \), polar constant \( (\Im) \), resonance constant \( (\Re) \), Hammet \( met\) constant \( (\sigma_m) \), Hammet \( para\) constant \( (\sigma_p) \), and aromatic group dipole \( (\mu) \). In addition, five input variables containing random data were included at each position to give a total of 72 input variables.
Chapter 3. Selective Descriptor Pruning

A second set of experimental data taken from the literature [Mathot et al., 1995; van der Graaf et al., 1999] consisted of adenosine A_1 receptor agonists and their protein binding values (fraction unbound, \( f_u \)) in rat. Each compound was represented by electronic, steric, and molecular surface volume descriptors, as well as 3D principal properties and calculated log \( P \) (clog \( P \)), all generated using commercially available computer programs. In addition to the 16 original descriptors a further 16 containing random data were generated in the present study and included to give 32 input variables in total.

Values for these data sets are given in Appendix A2.

### 3.2.6 Artificial Neural Network Model

#### 3.2.6.1 Model Parameters

The multilayer perceptron program used was BioActivNet 98c [AiMaze Pty Ltd, 1997]. Layers in the ANN consisted of one output neuron, between two and three hidden neurons, input neurons corresponding to the number of input variables, and an additional bias neuron in each layer. All neurons employed a sigmoidal transfer function during the feed-forward process and back-propagation was performed according to the generalised delta rule (Section 2.6.1).

For each ANN run there were two sets of momentum and learning-rates. The associated learning-rate switch was the number of cycles at which BioActivNet switched from using the first to the second set of momentum and learning-rate values.
Network parameters including learning-rate, momentum, learning-rate switch, number of training cycles, and error cutoff values were specific to each data set. Optimum values were determined by inspection after trial runs were conducted on each data set.

3.2.6.2 Model Training

The number of training cycles required was established by following the early stopping method to avoid overtraining. Overfitting of data was avoided by appropriate cross-validation [Tetko et al., 1995]. Training was performed in batches of ten runs after which weights and outputs were saved into a spreadsheet. For AS data sets trained networks were saved after each batch run and predictive ability was tested with new data different from that of the training set.

Learning-rate switch and maximum number of cycles were initially determined for data sets with only the output values and one meaningful input variable (X/Y data sets). Learning-rate switch and maximum cycles were then adjusted for each data set depending on the number of cycles needed for the ANN to converge.

Each ANN was initialised with all connection weights assigned random values between +0·10 and -0·10 [Swingler, 1996]. As repetitions were automated and performed in batches of ten there were a number of runs in which the ANN did not properly converge: these were identified through examination of the final weights matrices and excluded from the results presented.

3.2.6.3 Pruning of Descriptors

Selective pruning was performed at the end of successive training and cross-validation runs. Initially, large groups of meaningless descriptors were pruned
followed by smaller groups as the number of input variables decreased [Maddalena & Johnston, 1995].

For the benzodiazepine data set six input variables, one from each position, were removed at a time. For the adenosine A1 receptor agonist data set three pruning runs were conducted in which four input variables were removed, then three runs in which two input variables were removed. Subsequent to this only one input variable was pruned per run. Input variables included in the models with the highest cross-validation correlation were then classed as meaningful while the remainder were classed as meaningless.

3.2.7 Signal:noise and Rho

Signal:noise ratio (S:N) values were calculated for both AS and experimental data sets from the absolute mean weight values for meaningful and meaningless data thus (Equation 3-16):

\[
S : N = \frac{\bar{W}_{\text{meaningful}}}{\bar{W}_{\text{meaningless}}}
\]

Equation 3-16

For AS data sets \( W_{\text{meaningful}} \) were weights assigned to meaningful X input variable(s) and \( W_{\text{meaningless}} \) were weights assigned to \( R_j \) to \( R_m \) input variables, where \( m \) is the number of random variables included.

Values of \( \rho \) were determined according to Equation 3-15.
3.3 Results and Discussion

3.3.1 Relevance of Data

In some QSPR studies descriptor variables are related linearly to the target property. This is particularly the case where models are constructed using multilinear regression. An example in pharmacokinetics is the concentration-dependency of drug absorption. In such a case the rate of absorption can be defined as the first-order absorption rate constant, $k_a$, which is proportional to the concentration of the drug at the site of absorption. An increase in the concentration at the site of absorption will then cause a proportional and linear increase in the rate of absorption.

On the other hand, nonlinear pharmacokinetics can be exhibited when the relationship between two parameters does not depend on the concentration but instead depends on the size of the dose. For example, the dose-dependent change in the clearance of paracetamol has been observed in humans and animals. Following a toxic dose of paracetamol, the ability of the body to eliminate the drug from the blood is impaired due to hepatotoxicity and depletion of conjugation substrates. Hence, the clearance of paracetamol for a subsequent dose will then be reduced.

Time-dependency is another example of nonlinear pharmacokinetics. The metabolism of certain drugs such as carbamazepine can exhibit time-dependent pharmacokinetics because of autoinduction effects. Carbamazepine stimulates its own metabolism so that clearance is usually quicker in multiple dosing compared with clearance after a single dose.
Pharmacokinetic data such as the plasma concentration of a drug during multiple dosing is another example of time-dependent data. Administration of the first dose causes an increase in the plasma concentration over time which peaks and then decreases as the rate of elimination exceeds absorption. A trough concentration is reached prior to the next dose of the drug, following which the sequence of events is repeated. In a maintenance-dose schedule the plasma concentration profile can be roughly approximated by a sine curve.

Plasma concentration can also be altered by the circadian rhythm of the body. The circadian rhythm itself is a time-dependent phenomenon and has been shown to affect time to peak and peak height of indomethacin plasma concentrations in humans [Clench et al., 1981].

### 3.3.2 Signal:Noise Ratio

The S:N value represented the relative importance assigned to the meaningful variables. If S:N was greater than one then the importance of the meaningful variables as determined by the ANN would be greater than the average importance assigned to the meaningless variables. A S:N less than one would indicate that the importance of the meaningful descriptors was less than the average importance assigned to the meaningless variables. Values were calculated as averages so the significance of the S:N ratio for a particular model would also be affected by the total number of both meaningful and meaningless input variables. Error bars were not plotted in Figure 3-12 since mean squared error values were all below 0.045.
3.3.2.1  **Linear and Time-Dependent Data**

S:N for linear, noisy linear and time-dependent AS data sets did not drop below one (Figure 3-12). This indicated that the ANN was consistently able to assign high weight values to meaningful input variables and lower weight values to the meaningless input variables. In relation to the pruning process in experimental data sets, a high S:N allows variables with low weights to be removed from the model with the confidence that they simply contain meaningless information.

![Figure 3-12](image_url)

**Figure 3-12.** S:N for a) linear and noisy linear and b) time-dependent AS data.
Chapter 3. Selective Descriptor Pruning

As the number of meaningless variables decreased S:N increased considerably. When moving from a large number of inputs to a small number during pruning, at a certain point the ANN appeared to identify the meaningful variables. From that point onwards S:N increased substantially as most of the weight was assigned to those meaningful variables.

ANNs generally provide similar results in comparison with multilinear regression studies for linearly-related data. Hence, their ability to identify meaningful variables amid large numbers of meaningless variables when a linear relationship exists is not surprising. ANNs have also been successful in areas involving time-dependent data such as in financial forecasting. Multilinear regression models have difficulty handling such data. The human brain would easily be able to identify the relationship between the meaningful and target variables upon inspection of the time-dependent data used in this study. ANNs have this capability as well after sufficient training. This demonstrates their effectiveness as nonlinear modeling tools.

3.3.2.2 Nonlinear Data

The nonlinear AS data was relatively complex compared with the linear AS data. S:N decreased to below one with the addition of large numbers of meaningless input variables to the nonlinear data (Figure 3-13). Accordingly, the ANN assigned a low weight value to the meaningful variable and higher weight values to some other, meaningless random variables. It should be noted that there was only one meaningful variable in the nonlinear AS data set whilst in experimental data sets there are usually more than one meaningful or important descriptor variable. Additional numbers of meaningful variables would tend to sequester
more of the total weight for themselves thus leaving a smaller surplus for the other, less meaningful variables. Another important point is that $S:N$ was derived from the mean weight of the random variables. Hence, there would be a spread of weight values around that mean value. According to the selective pruning method the variables presenting the lowest weights would be removed next. As a result of the large total number of variables at this stage, the low-weighted variables would have weights below the mean $W_{\text{meaningless}}$ value, and almost certainly below $W_{\text{meaningful}}$ as well. Moreover, as the number of input variables was reduced through pruning (Figure 3-12 and Figure 3-13) $S:N$ would further increase. As a result, there would be less danger of mistakenly removing one of the meaningful variables as the number of input variables decreased.

![Figure 3-13](image)

**Figure 3-13.** $S:N$ for nonlinear and noisy nonlinear AS data.

$S:N$ was seen to decrease slightly after particular rounds of pruning (Figure 3-13). This indicated minor chance correlations between the meaningless descriptors and
Chapter 3. Selective Descriptor Pruning

the output space. The decreases would not affect pruning since for low numbers of meaningless variables S:N was far in excess of one, while for high numbers of input variables and where S:N was less than one the argument given above would again hold true.

3.3.2.3 Literature Experimental Data

The highest cross-validation results for the benzodiazepine data set were achieved with the model containing 12 input variables. Consequently those 12 were designated as meaningful while the remainder were considered to be meaningless. S:N remained above one as input variables were pruned from the system (Figure 3-14a). From the point where 24 inputs remained S:N increased sharply in accordance with the trends observed in the AS data sets (Figure 3-12 and Figure 3-13).

Peak cross-validation correlation for the adenosine A₁ receptor agonist data set was achieved with three input variables remaining. Therefore these three were designated as meaningful. Initial S:N was close to two and increased over the course of pruning (Figure 3-14b). Maximum S:N values were apparent when there were few meaningless input variables left in the system. Although maximum S:N values for the experimental data were considerably lower than those reached for the AS data, for the purpose of pruning the difference between S:N values of two and 20 is, in practice, inconsequential. Results for AS data and experimental data should be expected to differ since the former merely provides a model for the latter. The two experimental data sets used in the present study did, however, lend evidence to trends exhibited by the AS data sets over the course of simulated
selective pruning. This was demonstrated by the generally increasing nature of S:N which remained above one during pruning of the experimental data sets.

**Figure 3-14.** S:N for experimental a) benzodiazepine and b) adenosine A₁ receptor agonist data taken from the literature.
3.3.3 Model Predictions and Rho

Plots of predictive correlation against ρ as the number of input variables was reduced by pruning were generated. Five different models for each AS data set were constructed and each was tested with the same independent data.

3.3.3.1 Linear Predictions

There was a definite hyperbolic relationship between predictive correlation and ρ for the linear AS data (Figure 3-15a). Lower values of ρ corresponded to lower predictive correlation. Slightly increasing ρ resulted in a substantial increase in predictive correlation and higher ρ values corresponded to a prediction correlation approaching unity.

Figure 3-15. Prediction correlation for a) linear and b) nonlinear AS data (± SD, n = 4).

Correlations for the noisy linear AS data (Figure 3-15b) mirrored the values observed for the pure linear AS data. In both cases prediction correlation...
plateaued at a high level from $\rho$ greater than about 1.5. For the 20-pattern data set this represented accurate predictions even with up to nine meaningless input variables present. Taken in context of the simple linear relationship apparent between the meaningful variable and output space, this indicated the robustness of the ANN in identifying any underlying relationship amongst the data even amid large amounts of meaningless information. The similarity of results for both linear and noisy data leant further strength to this observation.

### 3.3.3.2 Nonlinear Predictions

In a similar manner to the linear AS data the predictive correlation for the pure nonlinear AS data (Figure 3-16) remained low for small values of $\rho$.

![Prediction correlation for nonlinear and noisy nonlinear AS data (± SD, $n = 4$).](image)

From a certain point prediction correlation increased dramatically and for $\rho$ greater than 2.7 the prediction correlation plateaued at values approaching unity.
Below a $\rho$ of 2.7 prediction correlation was negligible to the extent that the ANN was unable to make reliable predictions. This was further demonstrated by the large oscillations (error) in prediction correlations for low values of $\rho$. For both pure and noisy nonlinear AS data sets the critical $\rho$ value of 2.7 represented four meaningless input variables besides the single meaningful variable.

High prediction correlations could only be achieved with fewer meaningless input variables for nonlinear AS data when compared with the linear AS data. This indicated that the relatively complex quadratic relationship amongst the data was more difficult for the ANN to model. Nevertheless, high prediction correlations were achieved with substantial amounts of meaningless information present in the data. Again, results were similar for both pure and noisy AS data.

3.3.3.3 Time-Dependent Predictions

The trend of low prediction correlations corresponding to low values of $\rho$ up to a certain point was again repeated for the time-dependent AS data (Figure 3-17). Increasing $\rho$ resulted in a sharp rise in prediction correlation which plateaued for $\rho$ greater than 2.2. Peak prediction correlations were not as high as those for linear and nonlinear AS data seemingly due to the nature of the data itself. This was evident since changing ANN parameters both initially as well as during training failed to achieve significantly higher prediction correlations. Low prediction correlations occur when there is insufficient information contained in the descriptor variables. Therefore, even though the two meaningful input variables for the time-dependent AS data were adequate for training purposes they did not
contain adequate information for good predictions. Regardless of peak prediction correlation attained however, the present study did not aim to quantitatively explore ANN predictive performance. The goal was instead to identify trends during selective descriptor pruning in ANN studies. Indeed, the relationship between prediction correlation and $\rho$ showed strong similarity for all of linear, nonlinear and time-dependent AS data. Substantial error was seen at low values of $\rho$ for all AS data (Figure 3-15, Figure 3-16, and Figure 3-17). This was indicative of large uncertainty in those ANN models. Hence, in addition to the poor level of accuracy, care should be exercised at low values of $\rho$ due to the unreliability in precision. It should be noted, however, that precision noticeably improved as $\rho$ increased such that high precision was achieved for the high prediction correlation values where it was most important.

Figure 3-17. Prediction correlation for time-dependent AS data ($\pm$ SD, $n = 4$).
3.3.3.4 Predictions for Experimental Data

The benzodiazepine data set from Maddalena & Johnston (1995) may be considered to contain 12 meaningful and necessary variables as described earlier. For the benzodiazepine data set cross-validation correlations were low for low values of $\rho$ (Figure 3-18a).

![Figure 3-18](image)

**Figure 3-18.** Prediction correlation for a) benzodiazepine ($\pm$ SD, $n = 3$) and b) adenosine $A_1$ receptor agonist data sets ($\pm$ SD, $n = 4$).
Cross-validation correlation increased to maximal values with increasing \( \rho \) which was consistent with results obtained using AS data. Maximum cross-validation correlation was achieved at a \( \rho \) of 1.9 with pruning of further input variables causing a reduction in predictive performance.

**Table 3-5.** Best models of various QSAR and QSPR studies using ANNs and their associated value of \( \rho \).

<table>
<thead>
<tr>
<th>Study</th>
<th>Best ( r^a )</th>
<th>( \rho )</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plasma protein binding of ( \beta )-blockers</td>
<td>0.98(^c)</td>
<td>0.5</td>
<td>[Gobburu &amp; Shelver, 1995]</td>
</tr>
<tr>
<td>EC(_{50}) of capsaicin analogues, region B(^*)</td>
<td>0.945(^c)</td>
<td>0.9</td>
<td>[Hosseini et al., 1997]</td>
</tr>
<tr>
<td>Drug concentration at sampling site</td>
<td>&gt;0.99(^b)</td>
<td>1.4</td>
<td>[Gobburu &amp; Chen, 1996]</td>
</tr>
<tr>
<td>( \alpha_1 )-adrenoceptor affinity of arylpiperazines</td>
<td>0.991(^d)</td>
<td>1.5</td>
<td>[Lopez-Rodriguez et al., 2001]</td>
</tr>
<tr>
<td>Ligand binding IC(_{50}) of benzodiazepines</td>
<td>0.896(^c)</td>
<td>1.5</td>
<td>[Maddalena &amp; Johnston, 1995]</td>
</tr>
<tr>
<td>DHFR inhibition of pyrimidines</td>
<td>0.851(^c)</td>
<td>1.8</td>
<td>[So &amp; Richards, 1992]</td>
</tr>
<tr>
<td>Volume of distribution of ( \beta )-blockers</td>
<td>0.98(^c)</td>
<td>2.0</td>
<td>[Gobburu &amp; Shelver, 1995]</td>
</tr>
<tr>
<td>Anti-HIV activity of HEPT derivatives</td>
<td>0.959(^{b,d})</td>
<td>2.2</td>
<td>[Jalali-Heravi &amp; Parastar, 2000]</td>
</tr>
<tr>
<td>DHFR inhibition of triazines</td>
<td>0.897(^b)</td>
<td>3.4</td>
<td>[Andrea &amp; Kalayeh, 1991]</td>
</tr>
<tr>
<td>Mutagenicity of aromatic amines</td>
<td>0.804</td>
<td>7.5</td>
<td>[Villemin et al., 1993]</td>
</tr>
</tbody>
</table>

\(^a\)best correlation value obtained in study; \(^b\)testing set correlation; \(^c\)cross-validation correlation; \(^d\)training correlation.

In a similar fashion cross-validation correlations for the adenosine A\(_1\) receptor agonist data set were initially low for the lower values of \( \rho \) (Figure 3-18b). As meaningless descriptors were pruned from the system and \( \rho \) increased, so too did cross-validation correlation until only three meaningful input variables remained.
For the entirety of the study $\rho$ remained below one, with peak cross-validation correlation reached at a $\rho$ of 0.5. A truer evaluation of predictive ability could have been achieved with the use of test data independent of the training data. This was not possible, however, since the size of the current data sets did not permit division into separate training and test sets of compounds. Although not a true measure of predictive ability, cross-validation using the LOO method has been used frequently to give a fair indication of predictive performance of ANN models (Table 3-5).

### 3.3.4 The Effect of Rho

The ideal range proposed by Andrea & Kalayeh (1991) of $1.8 < \rho < 2.2$ has been accepted to apply broadly to QSAR studies involving ANNs as a rule-of-thumb [Lopez-Rodriguez et al., 2001]. However, their conclusion was based on experimental work carried out on one particular drug class and then applied to ANN studies in general. Closer examination of results presented by Andrea & Kalayeh (1991) revealed that peak cross-validation correlation was achieved for $\rho$ of 3.4 (Table 3-5). Furthermore, in “other unpublished work in which the surfaces are nearly linear, optimal predictions are obtained with models having $\rho > 0.5$.” Subsequent to the study by Andrea & Kalayeh (1991), So & Richards (1992) described the proposed guidelines for $\rho$ as being empirical and likely to depend upon the ANN application/study at hand. It can be seen that other studies achieved optimum results at different values of $\rho$ (Table 3-5). Implicitly, these data sets would contain different combinations of linear and nonlinear relationships amongst the various input and output spaces. From the results described in the
present study, optimum $\rho$ would vary according to the characteristic linearity and/or nonlinearity of individual data sets. The implications are that unless one knows precisely the extent of the linear and nonlinear relationships present within a given data set, basing the topology of an ANN on a pre-supposed value of $\rho$ is pointless and may even be detrimental to the QSAR/QSPR study at hand. Indeed, a frequently cited reason for the use of ANNs is that no prerequisite information concerning relationships amongst the data is needed for the construction of successful models [Aston & Wilding, 1992; Maddalena, 1996]. In addition, extraction of meaningful relationships from ANN models may prove difficult due to the uncertainty apparent in the weight matrices which do not consistently take the same values for the same final output [Aoyama et al., 1990]. High cross-validation correlations were achieved for $\rho$ greater than one and remained high as $\rho$ increased for experimental data (Figure 3-18b). This suggested that the relationships amongst the data cannot be classified solely as linear or nonlinear, and that good ANN performance can be achieved for a range of $\rho$ values. Thus, the ideal value of $\rho$ depends upon the nature of the data itself.

3.4 Conclusions

Selective pruning of descriptors from large data sets was performed. S:N for all linear and nonlinear AS data displayed a distinctly increasing trend as redundant variables were removed. This proved to be a robust method for optimising descriptor sets. A reduction in the number of input variables while maintaining other network parameters lead to an increase in $\rho$ during the optimisation process. As pruning was performed and $\rho$ increased, the prediction correlation increased
beyond a certain value of $\rho$. This value of $\rho$ varied between 0·5 and 2·7 for the different forms of data presented. These results indicated that the range of $1·8 < \rho < 2·2$ found by Andrea & Kaleyeh (1991) may not be generally applicable. Other observations summarised in Table 3-5 supported this conclusion.

Given that the relationship between input and output variables is not generally known, network architecture should not be constrained by an arbitrary range of $\rho$ as previously suggested by some authors. Rather, if a model is optimised by examining S:N using selective pruning and $\rho$ lies outside the range $1·8 < \rho < 2·2$, this should not necessarily be taken as evidence of a flaw in the architecture of the ANN model.
Chapter 4

QSPkRs for a Small Congeneric Set of Drugs: β-Adrenoceptor Antagonists
Chapter 4. QSPkRs for a Small Congeneric Set of Drugs: β-Adrenoceptor Antagonists

4.1 INTRODUCTION .................................................................................................................115
   4.1.1 β-Adrenoceptor Antagonists ..........................................................115
   4.1.2 QSPkR modeling..............................................................................116
   4.1.3 Study Aims ......................................................................................117

4.2 METHODS .........................................................................................................................118
   4.2.1 Literature Data ...............................................................................118
   4.2.2 Descriptor Generation ......................................................................119
      4.2.2.1 Topological descriptor calculation ........................................121
      4.2.2.2 Random Descriptor ................................................................123
   4.2.3 ANN Modeling ...................................................................................123
   4.2.4 Descriptor Selection ..........................................................................124
   4.2.5 Model Performance ............................................................................124
   4.2.6 Development of Regression Equations ...........................................125

4.3 RESULTS AND DISCUSSION .........................................................................................125
   4.3.1 Size of Data Set ................................................................................125
   4.3.2 Model Training ..................................................................................126
   4.3.3 Cross-Validation of Models ............................................................127
      4.3.3.1 Initial Models ............................................................................127
      4.3.3.2 Fluctuations during Cross-Validation .....................................130
   4.3.4 ANN Model Characteristics ............................................................131
      4.3.4.1 Efficiency Ratios ....................................................................131
      4.3.4.2 The Parameter ρ .....................................................................132
   4.3.5 Optimum Models ..............................................................................133
      4.3.5.1 Model Complexity ....................................................................133
      4.3.5.2 Descriptor Analysis ..................................................................134
      4.3.5.3 Validation Predictions and Experimental Values ...............136
   4.3.6 Further Tests and Comparisons .......................................................139
      4.3.6.1 Predictive Performance of Optimum Models .........................139
      4.3.6.2 ANN vs Multilinear Regression Prediction ..............................141
      4.3.6.3 Comparison of Cross-Validation with Other Reports ...........142

4.4 CONCLUSIONS ...............................................................................................................143
Chapter 4. QSPkRs for a Small Congeneric Set of Drugs: β-Adrenoceptor Antagonists

4.1 Introduction

An analysis of ANN model optimisation was presented in the preceding chapter. The present study employed the selective descriptor pruning method described to optimise QSPkR models for a series of β-adrenoceptor antagonists. The data set used in the present study was relatively small and contained structurally similar compounds. This was in order to establish the technique of theoretical descriptor generation and to demonstrate the application of selective descriptor pruning in QSPkR model construction for a relatively simple system. Separate models were constructed for six individual pharmacokinetic parameters. An analysis of the optimum models and their relation to each pharmacokinetic parameter was undertaken.

4.1.1 β-Adrenoceptor Antagonists

The β-adrenoceptor antagonists have been used widely for clinical treatment of congestive heart failure, ischemic heart disorders, hypertension, and certain arrhythmias. Their effect is more pronounced in tissues and states where sympathetic control is dominant. Hence, under normal physiological conditions their negative chronotropic and inotropic effects are moderate. Conversely, during stress or exercise when control by the sympathetic nervous system predominates they more effectively oppose the expected increase in heart rate. In a similar
manner, β-adrenoceptor antagonists generally do not cause a reduction in blood pressure in patients with normal blood pressure but do have considerable blood pressure-lowering effects in patients with hypertension [Mimran & Ducailar, 1988].

4.1.2 QSPkR modeling

As described in Section 1.3, conventional QSPkR analyses in general employ methods relating experimentally-derived properties such as tissue:blood partition coefficient and octanol:buffer partition coefficients to predict drug pharmacokinetic parameters. Experimental generation of this information is time and resource intensive, and has proven difficult because of the complex physiological processes involved in drug ADME and the nonlinear relationships present amongst drug data [Fouchecourt et al., 2001].

To reduce the complexity of modeling ADME processes, data sets containing structurally similar compounds can be used. Traditional approaches to structure-property relationship modeling have mainly involved congeneric sets of chemical compounds. The compounds consist of a common structural framework or scaffold around which functional groups and other structural substitutions are made. In this fashion the effect of individual substitutions on the target property can be closely examined. Another advantage of using a congeneric set of compounds is that there are less likely to be dramatic variations in certain properties such as diffusion and metabolism. Structurally dissimilar compounds can have substantially different molecular weights which can then affect their passage across biological membranes. Structurally similar compounds tend to
have molecular weight within a smaller range so differences in diffusion characteristics would be less pronounced than otherwise. In terms of metabolism, some enzyme classes are specific for particular structural motifs. The lower degree of structural variation exhibited by congeneric compounds can exclude metabolism by certain enzymes, thus decreasing the number of potential metabolic pathways.

Lower structural variation can have a number of effects on the values of theoretical descriptors. First, values of descriptors tend to be more correlated for congeneric series of drugs. Variations in structure for congeneric compounds occur at the substitution position while the scaffold remains constant. As a result, descriptors encoding information from part of or the entire scaffold contain the same information for the individual compounds. The selective pruning technique is useful in this case since removal of descriptors is partially based on their correlation with one another (Section 2.4). The second effect of reduced structural variation is that more specific information about structural effects are encoded in descriptors. Commonality of the scaffold ensures that changes in the target property of the model are caused by structural substitutions. The common scaffold also ensures that the structural substitutions themselves are directly encoded by the theoretical descriptors. Hence, descriptors in the optimum model should be closely related to the target pharmacokinetic parameter.

### 4.1.3 Study Aims

The aim of the present study was to generate a number of theoretical descriptors from the molecular graph. The descriptors and ANN modeling were used to
construct optimised QSPkRs in order to predict various pharmacokinetic parameters for a series of β-adrenoceptor antagonists.

4.2 Methods

4.2.1 Literature Data

The ANN technique develops data-driven models such that known information about drugs from empirical methods does not influence the system. Hence, unless some sort of indicator is presented to the ANN to distinguish between, for example, different experimental conditions used to determine each compound’s pharmacokinetic values, then consistency of the data becomes an important factor for ANN training.

Data for the present study was a congeneric series of β-adrenoceptor antagonists adapted from the literature [Hinderling et al., 1984a; Hinderling et al., 1984b]. Primary and secondary pharmacokinetic parameters were determined by Hinderling et al. (1984) after thoroughly surveying the literature and selecting appropriate values. Of the 14 compounds available 10 were used for training and cross-validation purposes. Complete pharmacokinetic data was unavailable for penbutol, practolol, oxprenolol, and sotalol. These compounds were reserved for testing purposes with the exception of penbutol which was discarded because available pharmacokinetic data lay outside the limits of the training set. Pharmacokinetic parameters investigated were renal and nonrenal clearance ($CL_R$ and $CL_{NR}$ respectively), volume of distribution at steady state ($V_{ss}$), volume of
distribution of unbound drug at steady state ($V_{u_{ss}}$), and fraction bound to plasma proteins ($f_b$).

**Table 4-6.** β-adrenoceptor antagonist data taken from the literature [Hinderling et al., 1984a; Hinderling et al., 1984b] and use in the present study.

<table>
<thead>
<tr>
<th>ID</th>
<th>Compound</th>
<th>Use</th>
<th>$f_b$</th>
<th>$V_{ss}$ (L)</th>
<th>$V_{u_{ss}}$ (L)</th>
<th>$CL_{NR}$ (mL·min$^{-1}$)</th>
<th>$CL_R$ (mL·min$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>01</td>
<td>acebutolol</td>
<td>training</td>
<td>0.26</td>
<td>93</td>
<td>126</td>
<td>475</td>
<td>295</td>
</tr>
<tr>
<td>02</td>
<td>alprenolol</td>
<td>training</td>
<td>0.76</td>
<td>75</td>
<td>313</td>
<td>435</td>
<td>18</td>
</tr>
<tr>
<td>03</td>
<td>atenolol</td>
<td>training</td>
<td>0.03</td>
<td>78</td>
<td>80</td>
<td>10</td>
<td>168</td>
</tr>
<tr>
<td>04</td>
<td>bufuralol</td>
<td>training</td>
<td>0.91</td>
<td>130</td>
<td>1440</td>
<td>535</td>
<td>4</td>
</tr>
<tr>
<td>05</td>
<td>metoprolol</td>
<td>training</td>
<td>0.08</td>
<td>223</td>
<td>242</td>
<td>701</td>
<td>93</td>
</tr>
<tr>
<td>06</td>
<td>nadolol</td>
<td>training</td>
<td>0.28</td>
<td>133</td>
<td>185</td>
<td>49</td>
<td>153</td>
</tr>
<tr>
<td>07</td>
<td>oxprenolol</td>
<td>test</td>
<td>0.92</td>
<td>-</td>
<td>-</td>
<td>174</td>
<td>22</td>
</tr>
<tr>
<td>08</td>
<td>penbutolol</td>
<td>excluded</td>
<td>0.98</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>09</td>
<td>pindolol</td>
<td>training</td>
<td>0.59</td>
<td>81</td>
<td>198</td>
<td>266</td>
<td>272</td>
</tr>
<tr>
<td>10</td>
<td>practolol</td>
<td>test</td>
<td>0.07</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>11</td>
<td>propranolol</td>
<td>training</td>
<td>0.93</td>
<td>137</td>
<td>1960</td>
<td>676</td>
<td>27</td>
</tr>
<tr>
<td>12</td>
<td>sotalol</td>
<td>test</td>
<td>0.00</td>
<td>95</td>
<td>95</td>
<td>-</td>
<td>159</td>
</tr>
<tr>
<td>13</td>
<td>timolol</td>
<td>training</td>
<td>0.60</td>
<td>100</td>
<td>250</td>
<td>524</td>
<td>70</td>
</tr>
<tr>
<td>14</td>
<td>tolamolol</td>
<td>training</td>
<td>0.91</td>
<td>148</td>
<td>1640</td>
<td>718</td>
<td>37</td>
</tr>
</tbody>
</table>

### 4.2.2 Descriptor Generation

Presentation of data containing adequately useful information to ANNs is the basis for construction of effective predictive models. A range of constitutional and 2D descriptors were generated to numerically encode meaningful features of each molecule (Table 4-7). The descriptors encoded hydrophobic, steric and electronic properties of each drug molecule. 3D descriptors were not used since the present study sought to avoid dependence on molecular conformation.

The following sections detail calculation of the majority of descriptors generated for the present study.
### Table 4-7. Molecular descriptors generated for β-adrenoceptor antagonist QSPkR.

<table>
<thead>
<tr>
<th>Descriptor Type</th>
<th>Symbol</th>
<th>No.</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Connectivity index differences&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0&lt;sup&gt;0&lt;/sup&gt;Δ&lt;sup&gt;ν&lt;/sup&gt;–4&lt;sup&gt;0&lt;/sup&gt;Δ&lt;sup&gt;ν&lt;/sup&gt;, 3&lt;sup&gt;Δ&lt;/sup&gt;Δ&lt;sup&gt;ν&lt;/sup&gt;–4&lt;sup&gt;Δ&lt;/sup&gt;Δ&lt;sup&gt;ν&lt;/sup&gt;, 4&lt;sup&gt;Δ&lt;/sup&gt;pc&lt;sup&gt;ν&lt;/sup&gt;</td>
<td>8</td>
<td>[Galvez et al., 1994b]</td>
</tr>
<tr>
<td>Connectivity index quotients&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0&lt;sup&gt;0&lt;/sup&gt;ζ&lt;sup&gt;ν&lt;/sup&gt;–4&lt;sup&gt;0&lt;/sup&gt;ζ&lt;sup&gt;ν&lt;/sup&gt;, 3&lt;sup&gt;ζ&lt;/sup&gt;ζ&lt;sup&gt;ν&lt;/sup&gt;–4&lt;sup&gt;ζ&lt;/sup&gt;ζ&lt;sup&gt;ν&lt;/sup&gt;, 4&lt;sup&gt;ζ&lt;/sup&gt;pc&lt;sup&gt;ν&lt;/sup&gt;</td>
<td>8</td>
<td>[Galvez et al., 1994b]</td>
</tr>
<tr>
<td>Charge indices&lt;sup&gt;c&lt;/sup&gt;</td>
<td>G1 – G5, G1&lt;sup&gt;v&lt;/sup&gt;– G5&lt;sup&gt;v&lt;/sup&gt;, J1 – J5, J1&lt;sup&gt;v&lt;/sup&gt;– J5&lt;sup&gt;v&lt;/sup&gt;</td>
<td>20</td>
<td>[Galvez et al., 1995]</td>
</tr>
<tr>
<td>Vertex counts&lt;sup&gt;d&lt;/sup&gt;</td>
<td>n, L, V3, V4</td>
<td>4</td>
<td>[Galvez et al., 1995]</td>
</tr>
<tr>
<td>Ramifications&lt;sup&gt;e&lt;/sup&gt;</td>
<td>Pr1, Pr2, Pr3</td>
<td>3</td>
<td>[Galvez et al., 1995]</td>
</tr>
<tr>
<td>Wiener number&lt;sup&gt;f&lt;/sup&gt;</td>
<td>W</td>
<td>1</td>
<td>[Galvez et al., 1995]</td>
</tr>
<tr>
<td>Molecular weight and derivatives&lt;sup&gt;g&lt;/sup&gt;</td>
<td>MW, ISRMW, ICRMW</td>
<td>3</td>
<td>[Jacobs, 1967; Herman &amp; Veng-Pedersen, 1994]</td>
</tr>
<tr>
<td>clog&lt;sup&gt;h&lt;/sup&gt; P and cross-products</td>
<td>At5, CDR, At5x, CDRx, clogPx</td>
<td>5</td>
<td>[Rekker &amp; De Kort, 1979; Viswanadhan et al., 1989]</td>
</tr>
<tr>
<td>Random</td>
<td>Ran</td>
<td>1</td>
<td>[Maddalena &amp; Johnston, 1995]</td>
</tr>
</tbody>
</table>

<sup>a</sup>chemical constitution of molecule; <sup>b</sup>linear combinations of connectivity indices derived from molecular graph; <sup>c</sup>derived from molecular graph and describe charge distribution; <sup>d</sup>counts of non-hydrogen atoms; <sup>e</sup>atom adjacency counts; <sup>f</sup>sum of topological distances in molecular graph; <sup>g</sup>diffusional characteristics; <sup>h</sup>log P calculated from molecular structure.

Values for clog P were determined using the PrologP 5.1 module in Pallas 2.0 [CompuDrug International, 1997]. In-house computer routines were written in Microsoft Visual Basic 6.0 to generate all other descriptors from the molecular graph. Atom and functional group counts [Kier & Hall, 1999], vertex counts...
[Galvez et al., 1995], and ramifications [Galvez et al., 1995] were totalled using simple counting routines.

### 4.2.2.1 Topological descriptor calculation

Kier and Hall connectivity (Equation 4-17) and valence connectivity indices (Equation 4-18) were calculated according to the following formulae [Kier & Hall, 1986]:

\[
\bar{m}_q = \sum_{k=1}^{K} \left( \prod_{a=1}^{n} \delta_a \right)^{-1/2} \\
\text{Equation 4-17}
\]

\[
\bar{m}^\prime q = \sum_{k=1}^{K} \left( \prod_{a=1}^{n} \delta^\prime a \right)^{-1/2} \\
\text{Equation 4-18}
\]

where \(k\) runs over all of the \(m\)th order subgraphs constituted by \(n\) atoms and \(n = m + 1\) for acyclic subgraphs. \(K\) is the total number of \(m\)th order subgraphs (see Section 1.5.2) present in the molecular graph and in the case of the path subgraphs \(K\) equals the \(m\)th order path count \(mP\). The product is over the simple vertex degrees \(\delta\) of all the vertices involved in each subgraph for nonvalence connectivity indices. Valence connectivity indices are calculated in the same manner but with the simple vertex index \(\delta\) being replaced by the valence vertex index \(\delta^\prime\). The subscript \(q\) refers to the type of molecular subgraph, and takes the value \(c\) for cluster, \(pc\) for path-cluster, \(ch\) for chain, and \(p\) for path. If \(q\) is omitted then the connectivity index is taken to be a path connectivity index.

Linear combinations of valence and nonvalence connectivity indices are thought to encode inductive and mesomeric effects related to the properties of electrons.
[Galvez et al., 1994b]. Differences (Equation 4-19) and quotients (Equation 4-20) were determined according to the following formulae:

\[ m\Delta_q = m\chi_q - m\chi_q^v \] \hspace{1cm} \text{Equation 4-19}

\[ m\zeta_q = \frac{m\chi_q}{m\chi_q^v} \] \hspace{1cm} \text{Equation 4-20}

Topological charge indices (Equation 4-21) and modified charge indices (Equation 4-22) encode charge transfers between pairs of atoms as well as the global charge transfer [Galvez et al., 1995]. They were defined as follows:

\[ kG = \sum_{i=1}^{N-1} CT_{ij} \delta(k, D_{ij}) \] \hspace{1cm} \text{Equation 4-21}

\[ kJ = \frac{kG}{(N - 1)} \] \hspace{1cm} \text{Equation 4-22}

where \( N \) is the number of vertices, \( \delta \) is Kronecker’s delta, and \( CT_{ij} = mij - mji \).

The term \( m \) stands for the elements of the \( M \) matrix: \( M = A \times D^* \) such that \( A \) is the adjacency matrix and \( D^* \) is the inverse square distance matrix. For both \( A \) and \( D^* \) the diagonal entries are given the value zero. In a similar manner to valence connectivity indices, valence charge indices were calculated by replacing the simple vertex index \( \delta \) with the valence vertex index \( \delta^v \) in the nonvalence formulae above.
4.2.2.2 **Random Descriptor**

An additional random descriptor [Maddalena, 1996] was also included as a quality control measure to monitor ANN performance. Examination of proportional weight assigned to the random descriptor indicated relative importance of other descriptors. The random descriptor was known to contain meaningless information. In some runs the random descriptor could have a higher mean absolute weight compared with other descriptors. That would indicate a relatively poor information content of those low-weighted descriptors. If the random descriptor was assigned high weight values that would then cast doubt on the reliability of the model. All descriptor values generated, excluding the random descriptor which was regenerated prior to each run, are given in Appendix A2.

4.2.3 **ANN Modeling**

The ANN program used was BioActivNet [AiMaze Pty Ltd, 1997]. Model construction and ANN parameters were similar to those described in Section 3.2.6.

ANN training followed the early-stopping rule to avoid overtraining [Tetko et al., 1995]. Initial number of training cycles were specific to each data set and determined by dummy-runs prior to model validation. Connection weights were initialised with random values between +0·10 and -0·10 and training runs were performed in batches of 10 to allow for models which did not converge. Results from unconverged runs were excluded from the final results presented.

Cross-validation was performed according to the LOO method to avoid overfitting [Tetko et al., 1995]. Further testing was undertaken with a testing set of data.
consisting of practolol, oxprenolol, and sotalol, which was independent from the training set. Prediction of the pharmacokinetic parameters for these compounds provided an unbiased indicator of real ANN predictive performance.

### 4.2.4 Descriptor Selection

Descriptor selection was performed according to the manual selective pruning technique described in Section 2.4.

### 4.2.5 Model Performance

Network performance was monitored by training correlation \( (r_t) \) and LOO cross-validation correlation \( (r_{cv}) \) values. Optimum models were determined according to an efficiency ratio (ER) [Maddalena, 1997] defined as the ratio of the latter to the former (Equation 4-23):

\[
ER = \frac{r_{cv}}{r_t}
\]

**Equation 4-23**

It has been found in ANN studies that \( r_{cv} \) is generally lower than \( r_t \), such that the efficiency for prediction of a test data set is lower than for prediction of the training set. Since \( r_t \) indicates how well the ANN manages the given data and \( r_{cv} \) is indicative of predictive performance then ER represents how efficient the ANN is at generalising.
4.2.6 Development of Regression Equations

Once optimum ANN models were determined the descriptors included were used to generate multilinear regression equations in Excel [Microsoft, 1997]. The purpose was to compare predictive performance of descriptors by linear and nonlinear methods.

4.3 Results and Discussion

4.3.1 Size of Data Set

ANNs represent learning tools which are distinctly different from standard statistical methods. Hence, ANNs are not necessarily bound by the same constraints that linear methods are. One important parameter in multilinear regression studies is the relationship between the number of experimental data points and optimisable parameters. A requirement for multilinear regression models is that the ratio of the former to the latter should be greater than a certain threshold. The required ratio for ANN models is not so straightforward, however, since the optimum value depends upon the nature of the data set itself [So & Richards, 1992]. It was found in the previous chapter that ρ, defined as the ratio of the number of compounds to the number of connections can vary greatly without compromising the results of an ANN model (see Chapter 3).
4.3.2 Model Training

All 73 descriptors generated were used to train the ANN, after which weights and correlations were analysed and pruning was implemented. One group of three and one group of four descriptors contained perfectly correlated information so two of the former and three of the latter were the first descriptors to be pruned. During initial pruning runs groups of descriptors were removed from the system [Maddalena & Johnston, 1995]. As the model improved pruning was limited to removal of one descriptor at a time. Training correlations were high for all data sets and remained above 0.990 (SD < 0.006) prior to optimum descriptor sets being reached. Over the course of pruning $r_t$ decreased slightly until the optimum model was reached, after which $r_t$ decreased substantially (Figure 4-19).

![Figure 4-19](image.png)

**Figure 4-19.** Training correlation for $CL_R$ over the course of pruning, representative of other pharmacokinetic parameters. Error bars not shown since SD values were all below 0.0004.
4.3.3 Cross-Validation of Models

4.3.3.1 Initial Models

Large numbers of input variables cause overfitting of data resulting in models with a poor ability to generalise. This was the case with initial models since all descriptors were included. Initial $r_{cv}$ values for $CL_R$, $CL_{NR}$, $V_{ss}$, and $f_b$ were all low as expected (Figure 4-20 and Figure 4-21), with $r_{cv}$ for $CL_R$ and $f_b$ close to zero. This indicated that predictions approximated random values.

![Graph](image1)

![Graph](image2)

**Figure 4-20.** β-adrenoeceptor antagonists: cross-validation correlation ($r_{cv}$) over the course of pruning for a) $CL_R$ and b) $CL_{NR}$. 

---

Chapter 4. β-adrenoeceptor antagonist QSPkRs

127
Figure 4-21. β-adrenoceptor antagonists: cross-validation correlation ($r_{cv}$) over the course of pruning for a) $V_{ss}$, b) $V_{u_{ss}}$, and c) $f_b$. 
Even though initial absolute $r_{cv}$ for $CL_{NR}$, and $Vu_{ss}$ appeared reasonably high, error in those predictions was large (Figure 4-20b and Figure 4-21b). This was demonstrated by large deviations from the observed experimental values (Figure 4-22). The deviations were particularly apparent for values at the extremity of the data range. In addition, SD values given by the error bars were for most points unacceptably high. Errors between replicates of cross-validation testing were also high (Figure 4-20 and Figure 4-21). Hence, initial models incorporating all 76 descriptors could not be considered useful in terms of accuracy. This was in contrast to improved accuracy after pruning was completed (Figure 4-24) which is discussed further in Section 4.3.5.3.

In addition to accuracy, precision of predictions during early pruning was poor. This was demonstrated by the high standard deviations in $r_{cv}$ with zero to 40 descriptors pruned out (Figure 4-20 and Figure 4-21). The standard deviations between $r_{cv}$ values noticeably improved during pruning as $r_{cv}$ increased. Subsequent to optimum models being achieved standard deviations increased again as $r_{cv}$ decreased. Thus, once accurate predictive models were being achieved then precision of predictions also improved.
Chapter 4. β-adrenoceptor antagonist QSPkRs

4.3.3.2 Fluctuations during Cross-Validation

Over the course of pruning $r_{cv}$ increased for all pharmacokinetic parameters. After a number of runs, in the early stages of pruning, $r_{cv}$ actually decreased (Figure 4-20 and Figure 4-21). This was most likely due to the method of pruning employed. Up to four descriptors which were considered meaningless were removed at a time. Even though individually the descriptors which were removed were meaningless, the group of descriptors as whole may have contained some useful information. Even so, many of the descriptors were correlated amongst

![Figure 4-22. Predicted vs observed experimental values for initial models for a) $CL_{NR}$ and b) $V_{u_{ss}}$, where the dashed line is the line of identity (± SD, $n = 5$).](image-url)
themselves so that removal would not have entirely eliminated their information content from the system [Basak et al., 2000a]. Fluctuations in \( r_{cv} \) during early pruning were not considered significant because useful predictive models were only achieved in the later stages of pruning. In these later stages smaller groups or individual descriptors only were removed. Care was taken not to eliminate important descriptors.

**4.3.4 ANN Model Characteristics**

Pruning served the purpose of reducing the complexity of the ANN. Hence, once optimum models were achieved there was little chance of overtraining influencing the results.

**4.3.4.1 Efficiency Ratios**

Each pharmacokinetic parameter studied was described by different optimum models. The ER scores for all optimum models were the highest achieved over the course of model construction and pruning. All training correlations were very high and close to unity. Therefore ER scores very closely mirrored \( r_{cv} \) values (Figure 4-23). This may not always be the case in ANN studies since \( r_t \) can decrease substantially over the course of pruning. ER can also indicate the reliability of data. If ER is greater than one then \( r_{cv} \) is greater than \( r_t \). Should this situation arise it may indicate the presence of inconsistent data or the presence of outliers. ER values for the optimum models are given in Table 4-8 while values for all other models are provided in Appendix A2.
Chapter 4. β-adrenoceptor antagonist QSPkRs

4.3.4.2 The Parameter $\rho$

The ratio, $\rho$, has previously been suggested to lie within the range $1.8 > \rho > 2.2$ for optimal results [Andrea & Kalayeh, 1991]. As described in Chapter 3 the more current and correct view is that $\rho$ is implementation dependent. The values of $\rho$ in the optimum models developed in the present study lay within the range $0.1 – 0.7$. Another study using the same experimental data as the present study obtained successful ANN models with $\rho$ having minimum and maximum values of 0.5 and 2.0 respectively [Gobburu & Shelver, 1995]. The combination of selective pruning and the nature of the data itself enabled the development of sound ANN
Chapter 4. β-adrenoceptor antagonist QSPkRs

models in the present study. This was further evidenced by the high \( r_{cv} \) values and predictive ability of the models developed which would not be the case if the models were unsound.

### 4.3.5 Optimum Models

With respect to the optimum models developed, it was found that varying numbers of neurons in the hidden layer of the ANN were required to best model the individual pharmacokinetic parameters (Table 4-8). The second finding was that optimum models contained distinct combinations of descriptors. The third finding was that, in most cases, optimum models contained a different final number of descriptors.

#### Table 4-8. Optimum ANN models for β-adrenoceptor antagonists attained after pruning.

<table>
<thead>
<tr>
<th>PK parameter(^a)</th>
<th>( r_{cv} \pm SD )</th>
<th>ER</th>
<th>Descriptors</th>
<th>Hidden neurons(^b)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( CL_R )</td>
<td>0.995 ± 0.003</td>
<td>0.996</td>
<td>112C, 131C, 221N, ( 4\zeta_v ), CDRx</td>
<td>2</td>
</tr>
<tr>
<td>( CL_{NR} )</td>
<td>0.995 ± 0.001</td>
<td>0.995</td>
<td>131C, 212N, 320O, C, Pr3, CDR, ( 0\zeta_v ), ( 3\Delta_v ), ( 4\zeta_{pc} ), ( 5\zeta_v ), G5, J1, L, ICRMW</td>
<td>6</td>
</tr>
<tr>
<td>( V_{ss} )</td>
<td>0.972 ± 0.004</td>
<td>0.973</td>
<td>320O, N, O, ( 1\zeta_v ), ( 4\zeta_{pc} )</td>
<td>5</td>
</tr>
<tr>
<td>( V_{us} )</td>
<td>0.993 ± 0.003</td>
<td>0.994</td>
<td>112C, 131C, 221N, ISRMW, ( 4\zeta_{pc} ), J1, V3, Pr1, Pr2</td>
<td>3</td>
</tr>
<tr>
<td>( f_b )</td>
<td>0.97 ± 0.01</td>
<td>0.97</td>
<td>113C, 131C, S, ( 3\zeta_v ), G3, Pr1, CDRx, clogPx</td>
<td>2</td>
</tr>
</tbody>
</table>

\( ^a \)pharmacokinetic parameter; \( ^b \)number of neurons in hidden layer of ANN.

#### 4.3.5.1 Model Complexity

The number of descriptors required to predict a given pharmacokinetic parameter may reflect the relative complexity of that particular parameter. Clearance is a measure of drug elimination from the body and occurs due to many physiological
Chapter 4. β-adrenoceptor antagonist QSPkRs

processes. As long as linear elimination processes are involved total body clearance can be expressed as the sum of $CL_R$ and $CL_{NR}$. In comparison, $CL_R$ is relatively simple and generally involves the processes of glomerular filtration and reabsorption. Active tubular secretion by non-specific cation transporters also plays a major role in renal clearance of β-adrenoceptor antagonists [Somogyi et al., 1996]. In contrast, $CL_{NR}$ of β-adrenoceptor antagonists is more complex since this class of drugs undergoes a wide variety of metabolic reactions primarily in the liver. This was reflected by the relative complexity of the optimum ANN model for each parameter: six hidden neurons and 13 descriptors were required for $CL_{NR}$ while only two hidden neurons and five descriptors for $CL_R$. High predictive performance was achieved for these pharmacokinetic parameters with very similar $r_{cv}$ values being recorded for both.

By definition, $V_{ss}$ is determined by a number of different physiological processes including protein binding. Therefore, modeling of $V_{ss}$ would necessitate the other physiological processes as well as protein binding to be taken into account. The optimum model describing protein binding contained eight descriptors and two hidden neurons. As described above, a model for $V_{ss}$ would be expected to be more complex than protein binding alone. In support of this, the optimum model for $V_{ss}$ contained greater numbers of descriptors and hidden neurons than the optimum model for protein binding (Table 4-8).

4.3.5.2 Descriptor Analysis

A number of descriptors remaining in optimum models described features of drugs generally known to relate to their physicochemical properties. Optimum models for $CL_R$, $CL_{NR}$, and $f_b$ all included clog $P$ descriptors which are closely
related to lipophilicity. Lipophilicity has been used to describe the dissociation constant $K_D$ between a drug-protein complex for a homologous series [Seydel & Schaper, 1981]. Hence, clog $P$ descriptors were expected in final models for $f_b$. Furthermore, only unbound drug is able to undergo glomerular filtration so clog $P$ by way of $f_b$ was expected to indirectly influence $CL_R$ prediction.

Molecular weight derivatives such as the inverse cube-root of molecular weight (ICRMW) and inverse square-root of molecular weight (ISRMW) have been shown to describe molar volume [Jacobs, 1967]. Molar volume can influence diffusion properties of compounds through biological membranes. Molecular weight derivatives were present in models for $CL_{NR}$ and $V_{us}$ which indicated that diffusion properties of drugs may be important for prediction of these pharmacokinetic parameters.

Connectivity indices up to the fourth order are known to encode various molecular properties including molecular density, branching and aromatic ring substitutions [Kier & Hall, 1986]. They have also recently been correlated with $f_b$ [Murcia-Soler et al., 2001]. Linear recombination of connectivity indices provide more useful information for prediction [Galvez et al., 1994b; Galvez et al., 1995] so it was reasonable to expect their inclusion in optimum models for $f_b$, $V_{ss}$, and $CL_R$.

Connectivity indices derive information from the hydrogen-depleted molecular graph where all vertices are considered equal. Charge indices rely on the valence state of graph vertices so encode information about the heterogeneity of a molecule. Different atoms can contribute in different ways to the properties exhibited by a molecule. For example, an aromatic ring composed of carbon
atoms provides a hydrophobic region which can promote π-stacking and alignment with other aromatic rings. This is in contrast to the positioning of an oxygen atom between two carbon atoms forming an ether group. Such an arrangement exhibits two lone-pairs of electrons on the oxygen which can then partially polarise the atom. This may then permit interaction with polar sites of other small molecules or proteins and thus affect whole-molecule behaviour.

All optimum models contained both charge indices and functional groups counts. This indicated the significance of structural considerations at an atomistic level in determining the pharmacokinetic behaviour of a compound.

A number of related pharmacokinetic parameters contained one or more of the same descriptors. For example, $4\zeta_{pc}v$ was in models for $V_{ss}$ and $V_{u,ss}$, and 131C was in models for $V_{u,ss}$ and $f_B$ (Table 4-8). This implies that some descriptors carry information useful for the prediction of a number of pharmacokinetic parameters.

Even so, the mechanistic relationship between pharmacokinetic parameters may not necessarily determine all the descriptors required to model a particular pharmacokinetic parameter. This is demonstrated by the variety of descriptors included in models for $V_{ss}$, $V_{u,ss}$, and $f_B$.

4.3.5.3 Validation Predictions and Experimental Values

Predicted pharmacokinetic parameter values were compared with experimental values (Figure 4-24). Most prediction data points showed strong agreement with experimental values, and in particular clear segregation of high and low values was seen for $V_{u,ss}$ (Figure 4-24d). Error at each point given by the SD of five runs was generally very low. This demonstrated the precision of the optimum models.
Some points displayed a more obviously higher SD than others especially for low values of $f_b$ (Figure 4-24e). This did not, however, detract significantly from overall accuracy of results presented.

Deviation from experimental values tended to be positive for low values and negative for high values (Figure 4-24a). Briefly, this may be attributed to the sigmoidal transfer function used in the ANN. Boundary limits on the transfer function were set to incorporate some part of the curved extremities of the sigmoid curve to allow for the nonlinear nature of the pharmacokinetic data being treated. Hence, for a more linear fit the limits could be adjusted to exclude more of curved extremities of the sigmoid curve.
Figure 4-24. Predicted vs experimental plots of optimum models for a) $CL_R$, b) $CL_{NR}$, c) $V_{ss}$, d) $V_{uss}$, and e) $f_b$, where the dashed line is the line of identity (± SD, $n = 5$).
4.3.6 Further Tests and Comparisons

True predictive performance of optimum ANN models was tested with the independent testing set. Comparisons were made with multilinear equations developed in the present study. Cross-validation results and predictions were compared with those obtained in other published studies.

4.3.6.1 Predictive Performance of Optimum Models

Predicted $CL_R$, $V_{ss}$, and $f_b$ for oxprenolol, sotalol, and practolol respectively agreed with experimental values (Table 4-9). Predictions also qualitatively indicated that $CL_R$ for sotalol and $f_b$ for oxprenolol were relatively high. Neither $V_{Us}$ of sotalol nor $CL_{NR}$ of oxprenolol were predicted well in the present study. However, predicted $CL_{NR}$ for oxprenolol was similar to that made by Gobburu & Shelver (1995) which was an expected result because of the complexity of nonrenal routes of metabolism for these drugs. It should also be noted in Table 4-9 that Hinderling et al. (1984) did not use independent test data but based predictions on regression equations which included oxprenolol in the original fitting. The ANN models constructed from theoretical descriptors in the present study were generally more accurate than the ANN models developed from physicochemical data [Gobburu & Shelver, 1995]. They were also more useful than the multilinear regression models from Hinderling et al. (1984) since they were validated against external data.
Table 4-9. Comparison of observed experimental with predicted pharmacokinetic parameters for test drugs.

<table>
<thead>
<tr>
<th>ID</th>
<th>Drug</th>
<th>Pharmacokinetic parameter</th>
<th>Observed</th>
<th>ANN(^a) ± SD</th>
<th>MLR(^b)</th>
<th>Gobburu(^c)</th>
<th>Hinderling(^d)</th>
</tr>
</thead>
<tbody>
<tr>
<td>07</td>
<td>oxprenolol</td>
<td>$CL_R$ (mL/min)</td>
<td>22</td>
<td>26 ± 3.5</td>
<td>44</td>
<td>3.7</td>
<td>43</td>
</tr>
<tr>
<td>07</td>
<td>oxprenolol</td>
<td>$CL_{NR}$ (mL/min)</td>
<td>174</td>
<td>603 ± 1.4</td>
<td>760</td>
<td>643</td>
<td>490</td>
</tr>
<tr>
<td>07</td>
<td>oxprenolol</td>
<td>$f_b$ (mL/min)</td>
<td>0.92</td>
<td>0.56 ± 0.01</td>
<td>-4.6</td>
<td>0.73</td>
<td>-</td>
</tr>
<tr>
<td>12</td>
<td>sotalol</td>
<td>$CL_R$ (mL/min)</td>
<td>159</td>
<td>287 ± 2.0</td>
<td>291</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>12</td>
<td>sotalol</td>
<td>$V_{ss}$ (L)</td>
<td>95</td>
<td>120 ± 0.5</td>
<td>41</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>12</td>
<td>sotalol</td>
<td>$V_{Uss}$ (L)</td>
<td>95</td>
<td>776 ± 14.6</td>
<td>1484</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>10</td>
<td>practolol</td>
<td>$f_b$ (L)</td>
<td>0.07</td>
<td>0.06 ± 0.01</td>
<td>-2.69</td>
<td>0.03</td>
<td>-</td>
</tr>
</tbody>
</table>

\(^a\)ANN predictions in the present study; \(^b\)multilinear regression predictions in the present study; \(^c\)values predicted by Gobburu and Shelver (1995); \(^d\)values predicted by Hinderling et al. (1984).
4.3.6.2 ANN vs Multilinear Regression Prediction

As a further comparison, multilinear regression equations based on the descriptor sets for the optimum ANN models were developed (Equation 4-24 to Equation 4-28). These equations were then used to predict pharmacokinetic parameters for the drugs in the test set (Table 4-9).

\[
CL_{NR} = -880.0 - 77.1(131C) - 315.0(212N) + 269.0(320O) - 71.1(C) - 1124.0(ICRMW) - 594.2(3\zeta_{v}) + 1529.4(3\Delta_{cv}) + 581.7(4\zeta_{pc_{v}}) - 104.3(5G) - 824.5(1J) + 30.7(n) + 104.4(Pr3) + 254.7(CDR) \tag{Equation 4-24}
\]

\[
CL_{R} = -105.7 - 21.8(112C) + 38.3(131C) + 179.8(221N) - 12.7(4\zeta_{cv}) - 8.9(CDRx) \tag{Equation 4-25}
\]

\[
f_{b} = 22.0 - 0.5(113C) + 0.1(131C) - 0.6(S) - 18.9(0\zeta_{v}) + 2.3(4\Delta_{cv}) + 0.5(3G^{v}) - 0.01(Pr1) - 0.2(CDR) \tag{Equation 4-26}
\]

\[
V_{ss} = -2.8 + 55.7(320O) - 38.9(N) + 25.6(O) + 664.3(1\zeta_{v}) - 429.2(4\zeta_{pc_{v}}) \tag{Equation 4-27}
\]

\[
V_{u_{ss}} = 8407.0 - 244.9(131C) - 332.0(221N) - 52044.5(ISRMW) - 3804.5(4\zeta_{pc_{v}}) + 1848.9(1J) + 236.4(V3) + 82.0(Pr1) + 121.1(Pr2) \tag{Equation 4-28}
\]

The fit of these multilinear regression equations to the training set were all very high \((r > 0.994)\). However, none of the pharmacokinetic values predicted for the independent test set were more accurate than those made using the ANN (Table 4-9). For the cases where both ANN and multilinear regression predictions were inaccurate, the multilinear regression equations produced results that were more so. Moreover, the majority of predictions made using the multilinear regression equations lay outside any meaningful bounds. It is the predictive ability of a model that makes it useful rather than training ability. Thus, for the purpose of the
current study, ANNs were found to be the superior modeling technique compared with multilinear regression.

**4.3.6.3 Comparison of Cross-Validation with Other Reports**

Another advantage of the present technique is that generation of descriptors requires knowledge only of the chemical structure of the drug, thus making experimental studies unnecessary. The two other published reports using this data for prediction of pharmacokinetic parameters [Hinderling et al., 1984a; Gobburu & Shelver, 1995] utilised methods based on experimental determination of physicochemical data for use as model inputs. These included reversed phase high-performance liquid chromatography for calculation of apparent octanol/buffer partition coefficient, the shake flask method for pK\textsubscript{a} and apparent octanol/buffer partition coefficient calculation, and experimentally-determined \( f_b \) as model inputs. Using only descriptors generated from drug structure the current study achieved \( r_{cv} \) values superior to those obtained from the other ANN and multilinear regression models using physicochemical descriptors in all cases except for one (Table 4-10). Even so, \( r_{cv} \) for \( f_b \) was high and predictions for independent drugs were comparable or superior to predictions made from models using experimentally-derived inputs (Table 4-9).

**Table 4-10.** Comparison of best cross-validation correlations.

<table>
<thead>
<tr>
<th>Pharmacokinetic parameter</th>
<th>Current study ± SD</th>
<th>Gobburu\textsuperscript{a}</th>
<th>Hinderling\textsuperscript{b}</th>
</tr>
</thead>
<tbody>
<tr>
<td>( CL_R )</td>
<td>0·995 ± 0·003</td>
<td>0·88</td>
<td>0·707</td>
</tr>
<tr>
<td>( CL_{NR} )</td>
<td>0·995 ± 0·001</td>
<td>0·91</td>
<td>0·757</td>
</tr>
<tr>
<td>( V_{ss} )</td>
<td>0·972 ± 0·004</td>
<td>0·96</td>
<td>-</td>
</tr>
<tr>
<td>( V_{Us} )</td>
<td>0·993 ± 0·003</td>
<td>0·98</td>
<td>0·911</td>
</tr>
<tr>
<td>( f_b )</td>
<td>0·97 ± 0·01</td>
<td>0·98</td>
<td>-</td>
</tr>
</tbody>
</table>

\textsuperscript{a}from Gobburu and Shelver (1995); \textsuperscript{b}from Hinderling et al. (1984).
4.4 Conclusions

Quality of data is the basis for production of successful predictive models. A set of structurally similar \( \beta \)-adrenoceptor antagonists and their experimentally determined pharmacokinetic parameters were used to develop a number of QSPkRs. Descriptive information about the drugs was derived solely from their chemical structure. The results demonstrated that useful information can be derived from molecular structure for a series of congeneric compounds. The selection of important descriptors was performed using manual selective pruning. Optimum descriptors selected in this manner allowed construction of better predictive models using ANNs when compared with multilinear regression equations. Model cross-validation values and predictions for independent test compounds were compared with other published results. These other studies employed experimentally derived physicochemical descriptors and/or ANN modeling. The combination of theoretically calculated descriptors and ANN modeling used in the present study achieved superior results. This method was successful in relating structural characteristics of a drug molecule to individual pharmacokinetic parameters.
Chapter 5

Multiple Pharmacokinetic Parameter Prediction for a Series of Cephalosporins
Chapter 5. Multiple Pharmacokinetic Parameter Prediction for a Series of Cephalosporins

5.1 INTRODUCTION ............................................................................................146
  5.1.1 Cephalosporin Antibiotics ...........................................................147
  5.1.2 Simultaneous Prediction of Drug Pharmacokinetics ..........147
  5.1.3 Study Aims .....................................................................149

5.2 METHODS ....................................................................................................149
  5.2.1 Cephalosporin Data ..............................................................149
  5.2.2 Descriptor Generation ...........................................................150
  5.2.3 ANN Model Construction .....................................................152

5.3 RESULTS AND DISCUSSION ..........................................................................155
  5.3.1 ANN Training ................................................................155
  5.3.2 Model Predictions ...............................................................156
    5.3.2.1 Optimum Model Selection ...........................................157
    5.3.2.2 Predictive Performance ..............................................159
  5.3.3 Descriptor Analysis ..............................................................161
    5.3.3.1 Linear Correlation of Descriptors...............................161
    5.3.3.2 Descriptor Sensitivities...............................................164
    5.3.3.3 Structure-Pharmacokinetic Relationships ..........165

5.4 CONCLUSION ...............................................................................................168
Chapter 5. Multiple Pharmacokinetic Parameter Prediction for a Series of Cephalosporins

5.1 Introduction

Most approaches in the field of structure-property relationship (SPR) analysis target only one property as the model objective. The present study undertook model development for the simultaneous prediction of a number of different pharmacokinetic parameters. Theoretical descriptors were generated for a set of structurally related commercial drugs and were used to construct a predictive ANN model. Work presented in this chapter has been published in the Journal of Pharmaceutical Sciences.

![Chemical structures](image)

**Figure 5-25.** Structure of a) the cephem nucleus on which the cephalosporins in the present study were based, and b) 7-aminocephalosporic acid.
5.1.1 Cephalosporin Antibiotics

Cephalosporins comprise a large number of related antibiotics commonly prescribed for the treatment of bacterial infections and as prophylactic agents. From cephalosporin C, one of the first of this class to be isolated, the active 7-aminocephalosporanic acid nucleus was determined (Figure 5-25b). This was then followed by modification of the parent compound by the addition of various side chains to form semi-synthetic compounds with improved antibacterial properties. It was found that modifications at position seven of the β-lactam ring alter the antibacterial properties of the compound. Substitutions at position three of the dihydrothiazine ring are associated with changes in the metabolism and pharmacokinetic properties [Petri Jnr, 2001]. Numerous cephalosporin derivatives exhibiting antibacterial activity have been developed by making such modifications and substitutions but the major cause for drug candidate failure has been unsuitable pharmacokinetics.

5.1.2 Simultaneous Prediction of Drug Pharmacokinetics

There is no single pharmacokinetic parameter that determines suitability of a drug for use in humans. Different pharmacokinetic parameters must be considered depending on the purpose of the drug in question as well as the physiological and pathological state of the patient. For example, too short a half life ($t_{1/2}$) or too rapid a clearance ($CL$) can mean that a pharmacologically effective compound may not reside in the body long enough to be therapeutically effective. On the other hand, drugs with longer residence times can be more prone to reaching toxic concentrations if patient plasma levels are not monitored. The fraction bound to plasma proteins ($f_{b}$) can also limit therapeutic activity because often only the free
unbound form of the drug is able to act at receptor sites or to cross biological membranes. Route of excretion may also play a role in certain disease states or physiological conditions. Should renal excretion be compromised such as in kidney disease or in the elderly, residence time and plasma concentrations of drugs excreted predominantly by this mechanism can reach dangerously high levels. As a result, predictive models for a single pharmacokinetic parameter alone do not give sufficient information regarding the potential clinical outcome. Therefore, instead of developing a model for one pharmacokinetic parameter only it would be prudent develop additional models to predict other important pharmacokinetic parameters. A single model able to predict multiple pharmacokinetic parameters simultaneously would be more time- and cost-effective than developing numerous models to perform the same task. The number of descriptor variables would also be reduced thus requiring less information than separate models would.

There are currently several commercially available computer programs for prediction of aqueous solubility, blood-brain barrier penetration, and absorption and permeability characteristics [Accelrys, 2001; Schrödinger, 2001] based on the chemical structure of a compound. In addition, a number of programs exist for the prediction of xenobiotic metabolism [Langowski & Long, 2002]. Some of these programs conveniently allow continuous updating of compound databases to tailor predictive results based on the scientist’s own data. Such programs, though, can be limited by requiring separate databases for the prediction of different physicochemical and pharmacokinetic properties. Furthermore, often separate
program modules need to be purchased and customised to provide worthwhile predictions for each physicochemical or pharmacokinetic property.

The majority of published QSPkR studies have constructed separate models for modeling of individual pharmacokinetic parameters [Herman & Veng-Pedersen, 1994; Gobburu & Shelver, 1995; van der Graaf et al., 1999]. In some cases the same physicochemical or theoretical descriptors are used but generally different information has been required for modeling of each pharmacokinetic parameter. One ANN study developed numerous models for the simultaneous prediction of human clearance and volume of distribution [Ritschel et al., 1995] for a set of structurally diverse drugs. The best models developed included combinations of both theoretical and physicochemical descriptors.

5.1.3 Study Aims

The present study aimed to develop a model for the simultaneous prediction of multiple pharmacokinetic parameters for a series of structurally related drug compounds. Information for the model would be non-experimental and derived from molecular structure alone.

5.2 Methods

5.2.1 Cephalosporin Data

The pharmacokinetic parameters of a series of 20 cephalosporins were collected from the literature. Compounds were divided into working (training and validation) and testing subsets (Table 5-11). Data for all compounds were taken
from studies employing intravenous administration of the drug, with the exception of cefaclor. Orally administered cefaclor has a bioavailability of >90% [Brumfitt & Hamilton-Miller, 1999]. ANNs are known to have robustness with managing noisy data [Hosseini et al., 1997] so oral pharmacokinetic data for cefaclor was accepted for model training instead. All compounds contained the cephem nucleus and were substituted at positions R₁ and R₂ (Figure 5-25a).

5.2.2 Descriptor Generation

A total of 133 theoretical descriptors were generated for each compound from molecular structure. These included constitutional, topological, chemical, geometrical, quantum chemical numbers, bulk properties, and solubility parameters (Table 5-12). Calculations were performed using Molecular Modeling Pro Demo 4.07 [ChemSW, 2001] and CAChe Project Leader 3.11 [Oxford Molecular, 2001]. Charge indices and linear combinations of topological descriptors were also calculated using in-house software (see Section 4.2.2). Many of the descriptors generated were 3D in nature so appropriate energy minimisation routines were employed as described in Section 2.3.
Table 5-11. Cephalosporin data set.

<table>
<thead>
<tr>
<th>Compound</th>
<th>Subset</th>
<th>$t_{1/2}$</th>
<th>$V_{ss}$</th>
<th>$CL$</th>
<th>$CL_R$</th>
<th>$f_e$</th>
<th>$f_b$</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>cefaclor</td>
<td>training</td>
<td>0.67</td>
<td>0.36</td>
<td>6.10</td>
<td>3.17</td>
<td>0.52</td>
<td>0.25</td>
<td>[Sides et al., 1988; Brumfitt &amp; Hamilton-Miller, 1999]</td>
</tr>
<tr>
<td>cefadroxil</td>
<td>training</td>
<td>1.20</td>
<td>0.24</td>
<td>2.90</td>
<td>2.70</td>
<td>0.93</td>
<td>0.20</td>
<td>[Welling et al., 1985]</td>
</tr>
<tr>
<td>cefamandole</td>
<td>validation</td>
<td>0.78</td>
<td>0.16</td>
<td>2.80</td>
<td>2.69</td>
<td>0.96</td>
<td>0.74</td>
<td>[Aziz et al., 1978]</td>
</tr>
<tr>
<td>cefazolin</td>
<td>training</td>
<td>1.80</td>
<td>0.14</td>
<td>0.95</td>
<td>0.76</td>
<td>0.80</td>
<td>0.89</td>
<td>[Scheld et al., 1981]</td>
</tr>
<tr>
<td>cefixime</td>
<td>training</td>
<td>3.00</td>
<td>0.3</td>
<td>1.30</td>
<td>0.53</td>
<td>0.41</td>
<td>0.67</td>
<td>[Brogden &amp; Campoli-Richards, 1989]</td>
</tr>
<tr>
<td>cefmetazole</td>
<td>training</td>
<td>1.50</td>
<td>0.18</td>
<td>1.45</td>
<td>1.16</td>
<td>0.80</td>
<td>0.70</td>
<td>[Ko et al., 1989]</td>
</tr>
<tr>
<td>cefonicid</td>
<td>validation</td>
<td>4.40</td>
<td>0.11</td>
<td>0.32</td>
<td>0.28</td>
<td>0.88</td>
<td>0.98</td>
<td>[Dudley et al., 1984]</td>
</tr>
<tr>
<td>cefoperazone</td>
<td>training</td>
<td>2.20</td>
<td>0.14</td>
<td>1.20</td>
<td>0.35</td>
<td>0.29</td>
<td>0.91</td>
<td>[Lode et al., 1980]</td>
</tr>
<tr>
<td>ceforanide</td>
<td>testing</td>
<td>2.60</td>
<td>0.14</td>
<td>0.26</td>
<td>0.22</td>
<td>0.84</td>
<td>0.81</td>
<td>[Estey et al., 1981]</td>
</tr>
<tr>
<td>cefotaxime</td>
<td>training</td>
<td>1.10</td>
<td>0.23</td>
<td>3.70</td>
<td>2.04</td>
<td>0.55</td>
<td>0.36</td>
<td>[Rodondi et al., 1989]</td>
</tr>
<tr>
<td>cefotetan</td>
<td>testing</td>
<td>3.60</td>
<td>0.14</td>
<td>0.23</td>
<td>0.15</td>
<td>0.67</td>
<td>0.85</td>
<td>[Martin et al., 1994]</td>
</tr>
<tr>
<td>cefpodoxime</td>
<td>training</td>
<td>2.30</td>
<td>0.46</td>
<td>2.40</td>
<td>1.94</td>
<td>0.81</td>
<td>0.27</td>
<td>[Chocas et al., 1993]</td>
</tr>
<tr>
<td>cefprozil</td>
<td>training</td>
<td>1.50</td>
<td>0.22</td>
<td>3.00</td>
<td>2.19</td>
<td>0.73</td>
<td>0.40</td>
<td>[Wiseman &amp; Benfield, 1993]</td>
</tr>
<tr>
<td>cefitzoxime</td>
<td>training</td>
<td>1.80</td>
<td>0.36</td>
<td>1.10</td>
<td>1.02</td>
<td>0.93</td>
<td>0.28</td>
<td>[Barriere &amp; Flaherty, 1984]</td>
</tr>
<tr>
<td>ceftriaxone</td>
<td>training</td>
<td>7.30</td>
<td>0.16</td>
<td>0.24</td>
<td>0.12</td>
<td>0.49</td>
<td>0.93</td>
<td>[Yuk et al., 1989]</td>
</tr>
<tr>
<td>cephalaxin</td>
<td>training</td>
<td>0.90</td>
<td>0.26</td>
<td>4.30</td>
<td>3.91</td>
<td>0.91</td>
<td>0.14</td>
<td>[Spyker et al., 1978]</td>
</tr>
<tr>
<td>cephalothin</td>
<td>training</td>
<td>0.57</td>
<td>0.26</td>
<td>6.70</td>
<td>3.48</td>
<td>0.52</td>
<td>0.71</td>
<td>[Bergan, 1987]</td>
</tr>
<tr>
<td>cephalpin</td>
<td>training</td>
<td>0.72</td>
<td>0.21</td>
<td>6.90</td>
<td>3.31</td>
<td>0.48</td>
<td>0.48</td>
<td>[Bergan, 1977]</td>
</tr>
<tr>
<td>cephradine</td>
<td>testing</td>
<td>0.90</td>
<td>0.46</td>
<td>4.80</td>
<td>4.13</td>
<td>0.86</td>
<td>0.14</td>
<td>[Schwinhammer et al., 1990]</td>
</tr>
<tr>
<td>loracarbef</td>
<td>testing</td>
<td>1.20</td>
<td>0.32</td>
<td>6.00</td>
<td>5.00</td>
<td>0.94</td>
<td>0.25</td>
<td>[Brogden &amp; McTavish, 1993; Sitar et al., 1994]</td>
</tr>
</tbody>
</table>

$t_{1/2}$ – half life (h); $V_{ss}$ – volume of distribution at steady state (L·kg$^{-1}$); $CL$ – clearance (mL·min$^{-1}$·kg$^{-1}$); $CL_R$ – renal clearance (mL·min$^{-1}$·kg$^{-1}$); $f_e$ – fraction excreted unchanged in the urine; $f_b$ – fraction bound to plasma proteins.
5.2.3 ANN Model Construction

All descriptors were included as inputs in the initial ANN model. Sensitivity analyses were performed to determine those inputs which had a significant effect on the model and those which did not. As opposed to using a sensitivity ratio (Section 2.4), the present study considered the absolute values of sensitivity for each descriptor. Descriptors with low absolute sensitivities were considered to be carrying redundant information and were pruned from the system. Pruning was automated so correlations between descriptors were not utilised for decision-making in this regard.

The supervised ANN model was constructed with NNmodel 1.512 [Neural Fusion, 1998]. NNmodel is a multilayer perceptron employing the generalised delta rule for back-propagation and weight adjustment (Section 2.6.1). Neurons in NNmodel contained a sigmoidal transfer function. All ANN models consisted of an input layer, one hidden layer with a variable number of neurons, an output layer containing six neurons, and a bias neuron connected to each layer. Output neurons corresponded to the six target pharmacokinetic parameters (Table 5-11).

ANN models were trained using all 14 compounds in the training set. Performance of each model over the course of training was monitored with the validation compounds which were not included during training. Training times and other parameters such as learning rate and momentum were chosen to optimise the training and predictive capabilities of the ANN models. Predictive performance of optimum models obtained after descriptor pruning was then evaluated using the independent testing set of compounds.
Numbers of neurons in the hidden layer were varied to enable more efficient model optimisation. Hidden neurons and, hence, network architecture determine the number of connections available in the model. All the information in an ANN resides in the strengths of these connections so the more connections that are available the more specific the information that can be stored by the model [Bucinski et al., 2000]. Initially, the large number of descriptor variables necessitated the inclusion of higher numbers of neurons but as pruning was performed the number of hidden neurons was able to be reduced.

Table 5-12. Calculated theoretical descriptors for cephalosporins.

<table>
<thead>
<tr>
<th>Descriptor Type</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Constitutional Descriptors</strong></td>
<td></td>
</tr>
<tr>
<td>⇒ percent weight: C, H, O, N, S, Cl, F</td>
<td></td>
</tr>
<tr>
<td>⇒ Functional group counts: amine, aldehyde, amide, carbonyl, carboxylate, cyano, ether, hydroxyl, methyl, methylene, nitro, nitroso, sulfide, sulfone, sulfoxide, thio</td>
<td></td>
</tr>
<tr>
<td><strong>Topological Descriptors</strong></td>
<td></td>
</tr>
<tr>
<td>⇒ connectivity indices: $^0\chi - ^4\chi$</td>
<td>[Gutman et al., 1999]</td>
</tr>
<tr>
<td>⇒ valence connectivity indices: $^0\chi^v - ^4\chi^v$</td>
<td>[Kier &amp; Hall, 1976]</td>
</tr>
<tr>
<td>⇒ connectivity index differences: $^0\Delta - ^4\Delta$</td>
<td>[Galvez et al., 1994b]</td>
</tr>
<tr>
<td>⇒ $^0\Delta^v - ^4\Delta^v, ^3\Delta^v - ^4\Delta^v, ^4\Delta^v_{pc}$</td>
<td></td>
</tr>
<tr>
<td>⇒ connectivity index quotients: $^0\zeta^v - ^4\zeta^v, ^3\zeta^v - ^4\zeta^v, ^4\zeta^v_{pc}$</td>
<td>[Galvez et al., 1995]</td>
</tr>
<tr>
<td>⇒ charge indices: $G1 - G5, G1^v - G5^v, J1 - J5, J1^v - J5^v$</td>
<td>[Randic, 2001]</td>
</tr>
<tr>
<td>⇒ Kier's topological shape indices: $\kappa^1 - \kappa^3$</td>
<td>[Trinajstic, 1992]</td>
</tr>
<tr>
<td>⇒ 3D Wiener number</td>
<td></td>
</tr>
</tbody>
</table>
## Chemical Descriptors

- parachor
- chemical properties: molecular mass, surface tension, polarisability, density
- partition coefficients: log $P$, log $D$, pKa, pKa$^0$

[Ahmad et al., 1975]
[Oxford Molecular, 2001]
[Reinhard & Drefahl, 1999]

## Geometrical Descriptors

solvent accessible surface, molar volume

[Oxford Molecular, 2001]

## Quantum Chemical Descriptors

- dipole moment, highest occupied molecular orbital (HOMO) and lowest unoccupied molecular orbital (LUMO) energies, dielectric energy, steric energy, heat of formation, total energy, minimum energy, electron affinity

[ChemSW, 2001]

## Bulk Properties

- molecular volume
- molar volume
- molecular weight, van der Waals volume, surface area, density, molecular length, width and depth

[Potts & Guy, 1993]
[Genty et al., 2001]
[Oxford Molecular, 2001]

## Solubility parameters

- Octanol-water partition coefficient: Q log $P$
- fragment addition log $P$
- atom based log $P$

[Bodor & Buchwald, 1998]
[Hansch, 1979]
[Viswanadhan et al., 1989]

- van Krevelen and Hansen’s solubility and 3D solubility parameters: dispersion, polarity and hydrogen bonding

[Crowley et al., 1966; Hansen, 1967]

- surface properties: polar surface area, surface tension, hydrophilic surface area and % hydrophilic surface area

[Ertl et al., 2000]

- hydrophilic-lipophilic balance: molecular weight and volumetric HLB

[ChemSW, 2001]

- mean water of hydration

[Meng & Carper, 2000]

- water solubilities: log $W$
- log molar water solubility ($log K_{ow}$), water solubility estimated from $log K_{ow}$ ($log S_w$)

[Klopman & Zhu, 2001]
[Cash & Clements, 1996]

- log molar olive oil - gas partition coefficient

[Klopman et al., 1997]

- molar refractivity, hydrogen bonding number

[Oxford Molecular, 2001]
5.3 Results and Discussion

Selection of relevant descriptors plays an important part in any QSPR study. A reduction in the number of descriptors by pruning led to a decrease in computing time. In addition, as the size of the ANN model was reduced its generalisation ability was expected to increase. Many of the descriptors were linearly correlated amongst themselves indicating that they contained a proportion of the same information. As such, redundant descriptors carrying correlated information were able to be removed without detriment to the model.

5.3.1 ANN Training

Inputs were pruned to 26, 18, 13 and finally 10 descriptors, with minimum absolute sensitivities of 1%, 2%, 3% and 4% respectively. Over the course of descriptor pruning ANN architecture changed. As the number of input neurons decreased different training correlations ($r_t$) were achieved (Table 5-13). The high training correlations overall demonstrated the ability of the ANN to process the input data. These training results also indicated that the ANN was able to correlate information presented by the input descriptors with the information presented in the output space to a high degree.

The variance of the data described by the trained model is given by the square of $r_t$. For the models presented in Table 5-13 both $r_t$ and variance values were all high. The lowest variance explained by a single model was 72% for $f_b$ by the ANN with 26 input variables and 10 hidden neurons. Even so, the variances explained by that particular model were greater than 92% for the remaining five
pharmacokinetic parameters. In all of the other models presented the variance explained was in excess of 81%.

Once the ANN was trained to an acceptable level, predictive performance for compounds external to the working set was tested.

### Table 5-13. ANN training performance achieved with different network architectures: values for a) training correlation, $r_t$, and b) variance, $r^2_t$.

<table>
<thead>
<tr>
<th>ANN$^a$</th>
<th>Training Performance</th>
<th>$t_{\frac{1}{2}}$</th>
<th>$V_{ss}$</th>
<th>$CL$</th>
<th>$CL_R$</th>
<th>$f_e$</th>
<th>$f_b$</th>
</tr>
</thead>
<tbody>
<tr>
<td>26-10-6</td>
<td>Correlation</td>
<td>0.968</td>
<td>0.99</td>
<td>0.971</td>
<td>0.960</td>
<td>0.97</td>
<td>0.85</td>
</tr>
<tr>
<td>18-9-6</td>
<td>Correlation</td>
<td>0.988</td>
<td>1.00</td>
<td>0.991</td>
<td>0.992</td>
<td>0.99</td>
<td>1.00</td>
</tr>
<tr>
<td>13-6-6</td>
<td>Correlation</td>
<td>0.973</td>
<td>0.92</td>
<td>0.981</td>
<td>0.985</td>
<td>0.97</td>
<td>0.94</td>
</tr>
<tr>
<td>10-6-6</td>
<td>Correlation</td>
<td>0.975</td>
<td>0.96</td>
<td>0.988</td>
<td>0.995</td>
<td>0.98</td>
<td>0.90</td>
</tr>
<tr>
<td>26-10-6</td>
<td>Variance</td>
<td>0.937</td>
<td>0.98</td>
<td>0.943</td>
<td>0.922</td>
<td>0.94</td>
<td>0.72</td>
</tr>
<tr>
<td>18-9-6</td>
<td>Variance</td>
<td>0.976</td>
<td>0.99</td>
<td>0.982</td>
<td>0.984</td>
<td>0.99</td>
<td>0.99</td>
</tr>
<tr>
<td>13-6-6</td>
<td>Variance</td>
<td>0.947</td>
<td>0.85</td>
<td>0.962</td>
<td>0.970</td>
<td>0.94</td>
<td>0.89</td>
</tr>
<tr>
<td>10-6-6</td>
<td>Variance</td>
<td>0.951</td>
<td>0.93</td>
<td>0.976</td>
<td>0.990</td>
<td>0.97</td>
<td>0.82</td>
</tr>
</tbody>
</table>

$^a$ANN architecture: inputs-hidden-outputs.

### 5.3.2 Model Predictions

Prediction correlations for the testing set were high for all architectures examined (Table 5-14). The simplest model contained six neurons in the hidden layer and required an input set of 10 descriptors. These 10 descriptors were all present in the next smallest model (13 descriptors), which in turn were all present in the model with 18 descriptors. The largest model with 26 descriptors contained all the descriptors in the other three models.
Table 5-14. ANN testing performance achieved with different network architectures: values for a) testing correlation, $r_{tes}$, and b) variance, $r_{tes}^2$.

<table>
<thead>
<tr>
<th>ANN</th>
<th>Testing Performance</th>
<th>$t_{1/2}$</th>
<th>$V_{ss}$</th>
<th>$CL$</th>
<th>$CL_R$</th>
<th>$f_e$</th>
<th>$f_b$</th>
</tr>
</thead>
<tbody>
<tr>
<td>26-10-6</td>
<td>Correlation</td>
<td>0.917</td>
<td>1.00</td>
<td>0.990</td>
<td>0.986</td>
<td>0.75</td>
<td>0.97</td>
</tr>
<tr>
<td>18-9-6</td>
<td>Correlation</td>
<td>0.946</td>
<td>0.51</td>
<td>0.999</td>
<td>0.994</td>
<td>0.77</td>
<td>0.85</td>
</tr>
<tr>
<td>13-6-6</td>
<td>Correlation</td>
<td>0.664</td>
<td>0.89</td>
<td>0.961</td>
<td>0.997</td>
<td>0.85</td>
<td>0.98</td>
</tr>
<tr>
<td>10-6-6</td>
<td>Correlation</td>
<td>0.927</td>
<td>0.96</td>
<td>0.647</td>
<td>0.926</td>
<td>0.92</td>
<td>0.99</td>
</tr>
<tr>
<td>26-10-6</td>
<td>Variance</td>
<td>0.842</td>
<td>0.99</td>
<td>0.981</td>
<td>0.972</td>
<td>0.56</td>
<td>0.95</td>
</tr>
<tr>
<td>18-9-6</td>
<td>Variance</td>
<td>0.896</td>
<td>0.26</td>
<td>0.998</td>
<td>0.987</td>
<td>0.59</td>
<td>0.72</td>
</tr>
<tr>
<td>13-6-6</td>
<td>Variance</td>
<td>0.440</td>
<td>0.79</td>
<td>0.924</td>
<td>0.994</td>
<td>0.72</td>
<td>0.95</td>
</tr>
<tr>
<td>10-6-6</td>
<td>Variance</td>
<td>0.859</td>
<td>0.91</td>
<td>0.419</td>
<td>0.857</td>
<td>0.84</td>
<td>0.98</td>
</tr>
</tbody>
</table>

*aANN architecture: inputs-hiddens-outputs.

5.3.2.1 Optimum Model Selection

Prediction correlations for the simplest model were not the highest achieved. Nevertheless, they were comparable with those gained using the other models (Table 5-14). Overall, testing correlations were lower than training correlations. This was to be expected since the supervised training methodology optimised the ANN to model relationships according to the training data. Therefore training correlations were expected be the highest achieved by the model compared with other correlations for validation or testing data.

The variances for the testing set explained by the models presented were correspondingly lower overall than the variances explained for the training set. In some cases the variance explained was below 50%, for example $t_{1/2}$ in the 13-6-6 model and $CL$ in the 10-6-6 model. However, these two models explained greater than 78% and 83% respectively of the variances for all of the five other pharmacokinetic parameters.
Chapter 5. Multiple PK Parameter Prediction

One approach in the development of models for the prediction of multiple pharmacokinetic parameters simultaneously could be to disregard outputs which appear to be modeled poorly. Thus, if during the optimisation process the variances of one or more pharmacokinetic parameters are explained poorly by the trained models then it could be ignored and optimisation would be directed to improve the model with respect to the other target outputs only. In this manner any predictions made by the final model for the ignored pharmacokinetic parameter would be relatively meaningless.

Another alternative for a poorly-handled pharmacokinetic parameter could be to remove it completely from the model. This would serve to reduce the complexity of the model since connections between the hidden and output layers would be removed as well. In addition, removal of a target variable would reduce the amount of information required by the model. Certain descriptors encoding information for the redundant output variable may no longer hold as large a sensitivity to the model overall and may then be subjected to pruning. These factors would all serve to decrease the complexity of the ANN model which may then lead to better generalisation.

As a general rule, complex models should be avoided in favour of simpler models. Since the smallest model with 10-6-6 architecture performed similarly to the other three ANN architectures it was selected as the optimum model and underwent further examination.
5.3.2.2 Predictive Performance

Correlation, $r$, can be defined as the linear dependence of one variable upon another and is independent of the unit of measurement. Hence, it is possible for two variables to be highly correlated and yet have vastly different absolute values. Of greater importance for prediction, then, is the magnitude of the pharmacokinetic values predicted by the model.

Predicted pharmacokinetic values for the validation set using the optimum model were compared with the observed experimental values taken from the literature (Figure 5-26). Target pharmacokinetic values in the present study were those for commercially available drugs. Hence, all pharmacokinetic values would be expected to lie within an acceptable range for human subjects. In relative terms, ceforanide and cefotetan have distinctly longer $t_{1/2}$ values than cephalexin and loracarbef. Predicted values of $t_{1/2}$ reflected this observation and absolute differences between predicted and observed values were acceptable (Figure 5-26a). Similar results were obtained for fraction bound to $f_b$ with predicted values closely mirroring the high or low observed values of the test compounds (Figure 5-26f). Observed values for volume of distribution and steady state ($V_{ss}$) and fraction excreted unchanged in the urine ($f_e$) were not so clearly divided into high or low values. However, predictions made by the optimum model were reasonably accurate in both relative and absolute terms (Figure 5-26b and Figure 5-26c). It is interesting to note that the predicted values of $CL$ and renal clearance ($CL_R$) for ceforanide were negative numbers (Figure 5-26c and Figure 5-26d). Drug clearance by the body is always positive. Therefore, an appropriate interpretation of the negative predicted pharmacokinetic values would be that the corresponding
experimental values lie close to zero. Ceforanide has a very low CL and hence low CL_R. The negatively predicted values were consistent with these observations.

![Graphs showing predicted values for cephalosporin test set](image)

**Figure 5-26.** Predicted values for cephalosporin test set for a) t_1/2, b) VUs, c) CL, d) CL_R, e) f_e, and e) f_b.
5.3.3 Descriptor Analysis

5.3.3.1 Linear Correlation of Descriptors

All descriptors were examined for linear correlations amongst themselves and with the target pharmacokinetic parameters. A high correlation was noted when the variance explained by one descriptor was greater than or equal to 50%. The majority of descriptors were highly correlated with at least one other descriptor. Many descriptors were highly correlated with numerous other descriptors.

Variances and correlations between descriptors in the optimum model and the target pharmacokinetic parameters were also examined (Table 5-15). Of the 45 pairs of descriptors there were eight pairs highly correlated with variances greater than 0.5 (Table 5-15 shown in bold print). Three of the descriptors (amine count, clog $K_{ow}$, and clog $P$) were considered independent from the other descriptors due to low correlations. Three descriptors were also correlated with two of the output parameters. $CL_R$ was correlated appreciably to the $1G^v$, $3G^v$, and LUMO, while $f_b$ was correlated also with $3G^v$ and LUMO. It was noted that $CL_R$ and $f_b$ were also correlated with each other.

Chance correlations may occur amongst data when dealing with multilinear regression problems in QSAR [Topliss & Edwards, 1979]. An analogy may be drawn with ANN QSPR studies but this has not yet been proven definitively. Mathematical examination of chance correlations makes the assumption that all descriptor variables are equally important in the model. The descriptor selection technique employed in the present QSPkR study sought to determine the most important descriptors from amongst a large number. Even when the optimum model was achieved the descriptors had varying importance within the model.
(Table 5-16). To test for chance correlations in a QSPkR, the model should be evaluated with an independent set of compounds. Predictive ability of the optimum model in the present study was assessed in this manner. Therefore, the optimum set identified in the present study represented important structural characteristics which collectively influence the pharmacokinetics of these cephalosporins. Intercorrelation between variables was a chance effect and it did not influence the reliability of the model.
Table 5-15. Squared correlation (variance) of optimum descriptor variables and observed pharmacokinetic parameters, with highly correlated values given in bold typeface.

<table>
<thead>
<tr>
<th>Variable</th>
<th>amides amines</th>
<th>1G⁺</th>
<th>3G⁺</th>
<th>4J⁻</th>
<th>4J⁺</th>
<th>clog K_{ow}</th>
<th>clog P</th>
<th>LUMO</th>
<th>molar volume</th>
<th>t(_{1/2})</th>
<th>V_{ss}</th>
<th>CL</th>
<th>CLR</th>
<th>f_e</th>
<th>f_b</th>
</tr>
</thead>
<tbody>
<tr>
<td>amides (^a)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>amines (^a)</td>
<td></td>
<td>0.03</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Charge index (1)</td>
<td></td>
<td>0.45</td>
<td>0.07</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Charge index (3)</td>
<td></td>
<td>0.54</td>
<td>0.07</td>
<td>0.70</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Charge index (4; J)(^b)</td>
<td></td>
<td>0.03</td>
<td>0.47</td>
<td>0.01</td>
<td>0.09</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Charge index (4; J)(^c)</td>
<td></td>
<td>0.05</td>
<td>0.20</td>
<td>0.13</td>
<td>0.23</td>
<td>0.75</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>clog K_{ow} (^d)</td>
<td></td>
<td>0.10</td>
<td>0.05</td>
<td>0.01</td>
<td>0.00</td>
<td>0.10</td>
<td>0.05</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>clog P (^e)</td>
<td></td>
<td>0.03</td>
<td>0.33</td>
<td>0.21</td>
<td>0.16</td>
<td>0.21</td>
<td>0.11</td>
<td>0.00</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LUMO (^f)</td>
<td></td>
<td>0.25</td>
<td>0.38</td>
<td>0.45</td>
<td>0.58</td>
<td>0.02</td>
<td>0.00</td>
<td>0.00</td>
<td>0.28</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>molar volume (^g)</td>
<td></td>
<td>0.38</td>
<td>0.02</td>
<td>0.52</td>
<td>0.51</td>
<td>0.05</td>
<td>0.25</td>
<td>0.24</td>
<td>0.01</td>
<td>0.18</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>t(_{1/2})</td>
<td></td>
<td>0.40</td>
<td>0.00</td>
<td>0.41</td>
<td>0.37</td>
<td>0.10</td>
<td>0.09</td>
<td>0.08</td>
<td>0.00</td>
<td>0.25</td>
<td>0.08</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>V_{ss}</td>
<td></td>
<td>0.15</td>
<td>0.22</td>
<td>0.14</td>
<td>0.28</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.12</td>
<td>0.49</td>
<td>0.07</td>
<td>0.13</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CL</td>
<td></td>
<td>0.18</td>
<td>0.03</td>
<td>0.41</td>
<td>0.50</td>
<td>0.03</td>
<td>0.09</td>
<td>0.18</td>
<td>0.16</td>
<td>0.41</td>
<td>0.07</td>
<td>0.47</td>
<td>0.22</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>CLR</td>
<td></td>
<td>0.24</td>
<td>0.09</td>
<td>0.62</td>
<td>0.54</td>
<td>0.01</td>
<td>0.04</td>
<td>0.08</td>
<td>0.28</td>
<td>0.57</td>
<td>0.10</td>
<td>0.48</td>
<td>0.30</td>
<td>0.82</td>
<td>1</td>
</tr>
<tr>
<td>f_e</td>
<td></td>
<td>0.24</td>
<td>0.00</td>
<td>0.27</td>
<td>0.09</td>
<td>0.00</td>
<td>0.00</td>
<td>0.27</td>
<td>0.05</td>
<td>0.04</td>
<td>0.18</td>
<td>0.04</td>
<td>0.02</td>
<td>0.01</td>
<td>0.07</td>
</tr>
<tr>
<td>f_b</td>
<td></td>
<td>0.26</td>
<td>0.31</td>
<td>0.44</td>
<td>0.50</td>
<td>0.02</td>
<td>0.00</td>
<td>0.01</td>
<td>0.20</td>
<td>0.80</td>
<td>0.17</td>
<td>0.35</td>
<td>0.63</td>
<td>0.36</td>
<td>0.55</td>
</tr>
</tbody>
</table>

\(^a\) functional group count; \(^b\) topological charge index; \(^c\) calculated log oil/water partition coefficient; \(^d\) calculated log n-octanol/water partition coefficient (fragmental); \(^e\) lowest unoccupied molecular orbital; \(^f\) calculated molar volume.
5.3.3.2 **Descriptor Sensitivities**

The sensitivity report of the final descriptor set in the optimum model (Table 5-16) demonstrated that the influence of a descriptor on one output pharmacokinetic parameter was not dependent upon its influence on the other output pharmacokinetic parameters.

<table>
<thead>
<tr>
<th>Descriptor</th>
<th>$t_\frac{1}{2}$</th>
<th>CL</th>
<th>$CL_R$</th>
<th>$f_e$</th>
<th>V</th>
<th>$f_b$</th>
</tr>
</thead>
<tbody>
<tr>
<td>amides$^a$</td>
<td>+0.091</td>
<td>+0.002</td>
<td>+0.013</td>
<td>+0.014</td>
<td>-0.038</td>
<td>+0.020</td>
</tr>
<tr>
<td>amines$^a$</td>
<td>+0.090</td>
<td>-0.066</td>
<td>-0.074</td>
<td>-0.017</td>
<td>+0.026</td>
<td>+0.035</td>
</tr>
<tr>
<td>charge index (1G)</td>
<td>+0.118</td>
<td>-0.088</td>
<td>-0.142</td>
<td>-0.089</td>
<td>-0.042</td>
<td>+0.158</td>
</tr>
<tr>
<td>charge index (3G)</td>
<td>+0.077</td>
<td>-0.160</td>
<td>-0.194</td>
<td>-0.045</td>
<td>+0.043</td>
<td>+0.103</td>
</tr>
<tr>
<td>charge index (4J)</td>
<td>-0.158</td>
<td>+0.187</td>
<td>+0.079</td>
<td>-0.163</td>
<td>-0.252</td>
<td>+0.049</td>
</tr>
<tr>
<td>charge index (4J)</td>
<td>+0.035</td>
<td>-0.081</td>
<td>+0.021</td>
<td>+0.162</td>
<td>+0.134</td>
<td>-0.116</td>
</tr>
<tr>
<td>clog $K_{ow}$</td>
<td>-0.012</td>
<td>+0.096</td>
<td>+0.003</td>
<td>-0.164</td>
<td>-0.028</td>
<td>+0.112</td>
</tr>
<tr>
<td>clog $P_d$</td>
<td>+0.020</td>
<td>+0.058</td>
<td>+0.031</td>
<td>-0.063</td>
<td>+0.080</td>
<td>-0.019</td>
</tr>
<tr>
<td>LUMO$^e$</td>
<td>-0.077</td>
<td>-0.014</td>
<td>+0.052</td>
<td>+0.107</td>
<td>+0.099</td>
<td>-0.169</td>
</tr>
<tr>
<td>molar volume$^f$</td>
<td>-0.116</td>
<td>+0.105</td>
<td>+0.194</td>
<td>+0.146</td>
<td>+0.020</td>
<td>-0.147</td>
</tr>
</tbody>
</table>

Table 5-16. Average sensitivity of descriptors for optimum model.

$^a$functional group count; $^b$topological charge index; $^c$calculated log oil/water partition coefficient; $^d$calculated log n-octanol/water partition coefficient (fragmental); $^e$lowest unoccupied molecular orbital; $^f$calculated molar volume.

For example, a simple pharmacokinetic relationship can be described between $CL$, $CL_R$ and $f_e$ where, because of the additivity of clearances in the body, $f_e$ is defined as the fraction of total $CL$ occurring via the kidneys [Rowland & Tozer, 1995d]. This relationship can be expressed as (Equation 5-29):

$$CL_R = CL \cdot f_e$$  \hspace{1cm} \text{Equation 5-29}$$

From the above equation it can be seen that concurrent increases in $CL$ and $f_e$ should cause a corresponding increase in $CL_R$. The charge index $4J'$ had positive sensitivities for both $CL$ and $f_e$, indicating that increasing the value of $4J'$ would cause an increase in both these values. The sensitivity of $4J'$ for $CL_R$ was negative.
indicating that an increase in $4J'$ would cause a decrease in $CL_R$. This observation directly contrasts with the relationship expressed in Equation 5-29.

In another example, LUMO energy had positive influences on $CL_R$, $f_v$, and $V_{ss}$, and negative influences on $t_{1/2}$, $CL$, and $f_b$. Increasing LUMO energy would therefore increase values of the former and decrease values of the latter. Mathematically, $t_{1/2}$, $V_{ss}$ and $CL$ can be related by the following (Equation 5-30) [Rowland & Tozer, 1995c]:

$$t_{1/2} = \ln\left(\frac{2 \cdot V}{CL}\right)$$

Equation 5-30

If descriptor sensitivities were dependent upon one another for all pharmacokinetic parameters then values would obey the relationship given in Equation 5-30. Thus, it would be expected that negative sensitivities for $t_{1/2}$ and $CL$ would correspond to a negative sensitivity for $V_{ss}$. On the contrary, this was not the case for the LUMO energy descriptor which had a positive sensitivity for $V_{ss}$.

Therefore, it was concluded that descriptors exert independent influences on each of the output pharmacokinetic parameters and are not necessarily bound by apparent mechanistic relationships.

5.3.3.3 **Structure-Pharmacokinetic Relationships**

The relative importance of descriptors can also be gauged by the sensitivity values such that higher absolute values indicate greater influence on the output. It can be seen that different descriptors were found to be important for different pharmacokinetic parameters (Table 5-16). Constitutional descriptors in the optimum model were amine and amide counts. Amine groups are basic in
character and amides are a resonance stabilised moiety. Amides contain a nitrogen with a single lone pair of electrons and an oxygen with two lone pairs of electrons. Functionally, the electronegative oxygen will tend to withdraw the electron pair from the nitrogen towards itself. Lone pairs can influence the electronic character of specific parts of the molecule itself as described, and can also play important roles in intermolecular interactions. Lone pairs are frequently involved in bond formation and can affect the dipole moment of the molecule.

Galvez charge indices [Galvez et al., 1994a] are derived from the molecular distance matrix in a similar fashion to Weiner indices. They have also been found to present useful information relating to the dipole moment. The inter- and intramolecular electronic effects related to dipole moment may influence metabolism or molecular reactivity and, owing to whole molecule polarity effects, may also influence membrane permeability. Topological indices generated from the distance matrix have been used for prediction of molecular properties such as boiling temperature and vaporisation enthalpy of organic molecules [Galvez et al., 1994a], and also pharmacological and pharmacokinetic properties of several drug classes [Galvez et al., 1995; Galvez et al., 1996]. The importance of charge indices for the prediction of $t_{1/2}$ is shown by their relatively high absolute sensitivity values (Table 5-16). In addition to coding charge properties these topological indices are algebraic descriptions of drug structure and, therefore, feature prominently in sensitivity analysis of all the other pharmacokinetic parameters predicted.

Further electronic properties of a molecule can be encoded by quantum mechanical descriptors. LUMO (lowest unoccupied molecular orbital) energy
represents the ability of a molecule to accept an electron during an interaction, thus encoding its reactivity as an electrophile. Lower LUMO energy levels correspond to better electrophilic properties and thus higher binding affinity. A high positive and high negative sensitivity for LUMO was found for $V_{ss}$ and $f_b$ respectively, indicating that binding played a significant role in determining the outcomes of these parameters.

Sensitivity values for molar volume were high for most pharmacokinetic parameters. This was especially true for $CL_R$ for which molar volume had the highest absolute value. Molar volume can account for certain diffusion processes such as membrane permeability. Diffusion across biological membranes can occur either across the cell membrane itself or in between cells forming the membrane. Molecular size and hence molar volume can limit this diffusion. Large molecules may be physically limited by their size and this can affect absorption of drugs administered nonparenterally. Reabsorption of molecules is generally a passive process so the importance of molar volume for the prediction of $CL_R$ is not surprising. In addition, molar volume has been related to aqueous solubility and the $n$-octanol/water partition coefficient [Reinhard & Drefahl, 1999] which are also important in renal excretory and reabsorption processes.

Classically, the logarithm of the octanol/water partition coefficient, log $P$, has been correlated with many physicochemical characteristics of molecules. Log $P$ is related to the lipophilic character of a molecule and has been used to describe passive membrane permeability and protein binding. Both calculated log $P$ (clog $P$) and calculated log $K_{ow}$ (clog $K_{ow}$) descriptors were included in the optimum model. The clog $K_{ow}$ descriptor had the highest absolute sensitivity for $f_e$. 

167
Lipophilic character of molecules can affect membrane permeability: low lipophilicity and hence high hydrophilicity will be disadvantageous to passive diffusion across a membrane. Hence, molecules displaying such characteristics often have difficulty crossing membranes. Higher lipophilicity values enable molecules to readily cross biological membranes. However, those with extremely high lipophilicity values may preferentially inhabit the membrane rather than passing through, thus slowing the diffusion process. Lipophilicity, then, could be expected to contribute to either $CL_R$ or $f_e$ because these are influenced by the degree of reabsorption of the drug back across a biological membrane. Although the relative contribution of lipophilicity for each parameter appears to be generally rather small, both clog $P$ and clog $K_{ow}$ were included in the final descriptor set from amongst a total of over 130 descriptors. Similarly, lipophilicity is known to influence protein binding. However, the absolute sensitivities for these two descriptors were not relatively high for prediction of $f_b$. The exact nature of the contribution of lipophilicity to drug pharmacokinetics is not clearly understood. Even so, the optimum model in the current study contained two descriptors related to lipophilicity which indicated its importance in physiological processes.

### 5.4 Conclusion

A model for the simultaneous prediction of drug $t_{1/2}$, $V_{ss}$, $CL$, $CL_R$, $f_e$ and $f_b$ was constructed for a series of cephalosporins. Predictive ability was validated with an independent set of compounds and predicted values agreed well with experimental literature values. Pruning of descriptors using sensitivity values was found to be a feasible alternative to manual selective pruning. Even though some descriptors in
the final set were mutually related this did not detract from overall model performance. Analysis of descriptor sensitivities demonstrated the deterministic nature of ANN modeling as distinct from physiologically based or mechanistic modeling. The major advantage of the method described in the present study was that a broad pharmacokinetic profile for a compound was able to be predicted from its molecular structure alone. This was accomplished with the use of a single optimised ANN model instead of separate models for each pharmacokinetic parameter.
Chapter 6

Bioavailability Prediction from Molecular Structure for a Diverse Series of Drugs
Chapter 6. Bioavailability Prediction from Molecular Structure for a Diverse Series of Drugs

6.1 INTRODUCTION .................................................................................................................. 172
6.1.1 Bioavailability ................................................................................................................. 172
6.1.2 Study Aims ....................................................................................................................... 173

6.2 METHODS ............................................................................................................................ 173
6.2.1 Descriptor Generation ...................................................................................................... 173
6.2.2 Drug Data Set ..................................................................................................................... 173
6.2.3 Input Variable Selection ................................................................................................... 180
6.2.4 Network Construction ....................................................................................................... 180
   6.2.4.1 Modified Efficiency Ratio ......................................................................................... 181
6.2.5 Stepwise Regression Modeling ....................................................................................... 182
   6.2.5.1 Regression Data ........................................................................................................ 182

6.3 RESULTS AND DISCUSSION – ANN MODEL .................................................................. 182
6.3.1 Descriptor Pruning .......................................................................................................... 182
6.3.2 Statistical Analysis ......................................................................................................... 183
6.3.3 Descriptor Analysis ........................................................................................................ 186
6.3.4 ANN Model Performance ............................................................................................... 195
6.3.5 Independent Predictions ................................................................................................. 198

6.4 RESULTS AND DISCUSSION – SWR MODEL .................................................................. 201
6.4.1 Model Construction and Performance ........................................................................... 202
6.4.2 Independent Predictions ................................................................................................. 203
6.4.3 Descriptor Analysis ........................................................................................................ 204
6.4.4 Comparison of ANN and SWR Models ......................................................................... 208

6.5 CONCLUSION ....................................................................................................................... 209
Chapter 6. Bioavailability Prediction from Molecular Structure for a Diverse Series of Drugs

6.1 Introduction

The previous two chapters dealt with structure-pharmacokinetic relationships derived for structurally related compounds. A quantitative structure-bioavailability relationship is described in the present study. A large, structurally diverse data set was used to construct a model able to screen potential drugs for suitable bioavailability. In addition to presenting a validated ANN model, another structure-bioavailability relationship was developed using multilinear regression. This provides a useful comparison between linear- and nonlinear-based models using a large non-congeneric data set. The ANN work presented in this chapter has been accepted for publication in Pharmaceutical Research. The stepwise regression work has been accepted for publication in Analytica Chimica Acta.

6.1.1 Bioavailability

Strategic decisions that affect bioavailability are considerably important in drug development and may cause a delay in new drug approval. Since the majority of new drugs are intended to be administered orally, the ability of a new drug to have good bioavailability is imperative. The major goal in pharmaceutical product development is for a drug to reach the site of action at a concentration sufficient to produce a therapeutic effect. Poor bioavailability can mean that a pharmacologically active compound may not be a clinically effective compound if
administered orally. Prediction of bioavailability would be advantageous to the pharmaceutical industry since it would provide an indication of whether a compound would be suitable for delivery as an oral formulation early in the development process.

6.1.2 Study Aims

The purpose of the present work was to construct a QSPkR for bioavailability using theoretical descriptors and ANNs. The model thus developed should be useful for virtual screening of structurally unrelated chemical compounds.

6.2 Methods

6.2.1 Descriptor Generation

Descriptors for the present study were similar to those calculated in the previous chapter (Section 5.2.2). A full list of the descriptors used and their values are given in Appendix A2. Descriptors were generated from drug structures according to the procedure outlined in Section 2.3.

6.2.2 Drug Data Set

The set of 167 structurally different compounds and their experimentally derived bioavailability values (%) used in this study were collected from the literature (Table 6-17). Screening of the literature for appropriate data was essential. Many early studies incorrectly considered absorption as oral bioavailability. Other studies published bioequivalence data where bioavailability of one formulation was compared with bioavailability of other established formulations. Values thus
obtained were expressed as an apparent bioavailability equal to the ratio of the former to the latter. These were of little value since the present study sought to develop quantitative and not relative relationships between structure and bioavailability. Quantitative values from more than one reference were sought where additional information regarding experimental bioavailability was required.

Prior to training, the data was divided randomly into three separate subsets: training (137 compounds), validation (15 compounds) and testing (15 compounds). Bioavailability values for the validation and testing subsets thus divided were examined statistically to ensure adequate representation of the training set ($p=0.91$).

Table 6-17. Drug and bioavailability data, including ANN predicted values.

<table>
<thead>
<tr>
<th>Drug</th>
<th>Use</th>
<th>Range</th>
<th>Target</th>
<th>Predicted</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>acyclovir</td>
<td>tra</td>
<td>15-30</td>
<td>23</td>
<td>44 ± 1.3</td>
<td>[Bras et al., 2001]</td>
</tr>
<tr>
<td>alendronate</td>
<td>tra</td>
<td>0.59-0.76</td>
<td>0.7</td>
<td>-6 ± 0.1</td>
<td>[Gertz et al., 1995]</td>
</tr>
<tr>
<td>allopurinol</td>
<td>tra</td>
<td>30-68</td>
<td>49</td>
<td>43 ± 0.2</td>
<td>[Appelbaum et al., 1982; Murrell &amp; Rapeport, 1986]</td>
</tr>
<tr>
<td>amantadine</td>
<td>tra</td>
<td>90</td>
<td>90</td>
<td>72 ± 0.3</td>
<td>[Aoki &amp; Sitar, 1988]</td>
</tr>
<tr>
<td>amiloride</td>
<td>tra</td>
<td>50</td>
<td>50</td>
<td>37 ± 0.5</td>
<td>[Grayson et al., 1971]</td>
</tr>
<tr>
<td>aminoglutethimide</td>
<td>tra</td>
<td>90</td>
<td>90</td>
<td>63 ± 0.1</td>
<td>[Santen &amp; Misbin, 1981]</td>
</tr>
<tr>
<td>amiodarone</td>
<td>tra</td>
<td>50</td>
<td>50</td>
<td>49 ± 0.5</td>
<td>[Popiliev, 1985; Fausa &amp; Rasmussen, 1986]</td>
</tr>
<tr>
<td>amitriptiline</td>
<td>tra</td>
<td>30-60</td>
<td>45</td>
<td>41 ± 0.2</td>
<td>[Schulz et al., 1983]</td>
</tr>
<tr>
<td>amlodipine</td>
<td>tra</td>
<td>60-64</td>
<td>62</td>
<td>52 ± 0.6</td>
<td>[Laine et al., 1997]</td>
</tr>
<tr>
<td>amoxicillin</td>
<td>tra</td>
<td>83-100</td>
<td>92</td>
<td>67 ± 0.7</td>
<td>[Jones &amp; Hill, 1974]</td>
</tr>
<tr>
<td>ampicillin</td>
<td>tra</td>
<td>25-75</td>
<td>50</td>
<td>66 ± 0.4</td>
<td>[Triggs et al., 1980]</td>
</tr>
<tr>
<td>aspirin</td>
<td>tra</td>
<td>65-71</td>
<td>68</td>
<td>81 ± 0.3</td>
<td>[Bochner &amp; Lloyd, 1995; Muir et al., 1997b]</td>
</tr>
<tr>
<td>atenolol</td>
<td>tra</td>
<td>45-55</td>
<td>50</td>
<td>63 ± 0.1</td>
<td>[Buck et al., 1989]</td>
</tr>
<tr>
<td>atorvastatin</td>
<td>tra</td>
<td>14</td>
<td>14</td>
<td>23 ± 0.2</td>
<td>[Gibson et al., 1996]</td>
</tr>
<tr>
<td>atropine</td>
<td>tra</td>
<td>50</td>
<td>50</td>
<td>56 ± 0.1</td>
<td>[Ellinwood et al., 1990]</td>
</tr>
<tr>
<td>azathioprine</td>
<td>tra</td>
<td>80</td>
<td>80</td>
<td>66 ± 0.1</td>
<td>[Tsutsumi et al., 1982]</td>
</tr>
<tr>
<td>baclofen</td>
<td>tra</td>
<td>70</td>
<td>70</td>
<td>81 ± 0.4</td>
<td>[Wuis et al., 1989]</td>
</tr>
<tr>
<td>bendrofluazide</td>
<td>tra</td>
<td>90</td>
<td>90</td>
<td>81 ± 0.1</td>
<td>[McAinsh et al., 1981]</td>
</tr>
<tr>
<td>Drug</td>
<td>Use</td>
<td>Range</td>
<td>Target</td>
<td>Predicted ± SD</td>
<td>Reference</td>
</tr>
<tr>
<td>-------------------</td>
<td>-----</td>
<td>-------</td>
<td>--------</td>
<td>---------------</td>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>betamethasone</td>
<td>tra</td>
<td>70</td>
<td>70</td>
<td>76 ± 0.4</td>
<td>[Petersen et al., 1983]</td>
</tr>
<tr>
<td>bromocriptine</td>
<td>tra</td>
<td>5-10</td>
<td>10</td>
<td>25 ± 0.3</td>
<td>[de Groot et al., 1998]</td>
</tr>
<tr>
<td>bumetanide</td>
<td>tra</td>
<td>80-100</td>
<td>90</td>
<td>78 ± 0.3</td>
<td>[Ward &amp; Heel, 1984]</td>
</tr>
<tr>
<td>captopril</td>
<td>tra</td>
<td>60-75</td>
<td>68</td>
<td>78 ± 0.3</td>
<td>[Mäntylä et al., 1984]</td>
</tr>
<tr>
<td>carbamazepine</td>
<td>tra</td>
<td>60-85</td>
<td>72</td>
<td>77 ± 0.1</td>
<td>[Liu &amp; Delgado, 1994]</td>
</tr>
<tr>
<td>cephalaxin</td>
<td>val</td>
<td>81-99</td>
<td>90</td>
<td>71 ± 0.2</td>
<td>[Bergan et al., 1970]</td>
</tr>
<tr>
<td>chloramphenicol</td>
<td>tra</td>
<td>75-90</td>
<td>83</td>
<td>80 ± 0.5</td>
<td>[Mulhall &amp; de Louvois, 1985]</td>
</tr>
<tr>
<td>chloridiazepoxide</td>
<td>tra</td>
<td>100</td>
<td>100</td>
<td>87 ± 0.2</td>
<td>[Minder, 1989]</td>
</tr>
<tr>
<td>chlorothiazide</td>
<td>tra</td>
<td>8-20</td>
<td>20</td>
<td>43 ± 0.6</td>
<td>[Adebayo &amp; Mabadeje, 1985]</td>
</tr>
<tr>
<td>chlortalidone</td>
<td>tra</td>
<td>65-75</td>
<td>70</td>
<td>74 ± 0.3</td>
<td>[Farina et al., 1985]</td>
</tr>
<tr>
<td>cimetidine</td>
<td>tra</td>
<td>60-70</td>
<td>65</td>
<td>62 ± 0.1</td>
<td>[Somogyi &amp; Gugler, 1983; Lesko et al., 2002]</td>
</tr>
<tr>
<td>ciprofloxacin</td>
<td>tra</td>
<td>50-70</td>
<td>60</td>
<td>66 ± 0.5</td>
<td>[Takamatsu et al., 1987; Campoli-Richards et al., 1988]</td>
</tr>
<tr>
<td>clindamycin</td>
<td>tra</td>
<td>50</td>
<td>50</td>
<td>70 ± 0.6</td>
<td>[Liliemark, 1997]</td>
</tr>
<tr>
<td>clonazepam</td>
<td>tra</td>
<td>1</td>
<td>1</td>
<td>28 ± 0.7</td>
<td>[Yakatan et al., 1982]</td>
</tr>
<tr>
<td>clonazepam</td>
<td>tra</td>
<td>95</td>
<td>95</td>
<td>103 ± 0.3</td>
<td>[Naito et al., 1987]</td>
</tr>
<tr>
<td>clonidine</td>
<td>tes</td>
<td>75-95</td>
<td>83</td>
<td>69 ± 0.4</td>
<td>[Manhem et al., 1982]</td>
</tr>
<tr>
<td>cloxacillin</td>
<td>tra</td>
<td>40-50</td>
<td>45</td>
<td>76 ± 0.8</td>
<td>[Nergelius et al., 1997]</td>
</tr>
<tr>
<td>cromoglycate</td>
<td>tra</td>
<td>2-5</td>
<td>3</td>
<td>46 ± 0.1</td>
<td>[Walker et al., 1972]</td>
</tr>
<tr>
<td>cyclophosphamide</td>
<td>tra</td>
<td>&gt;75</td>
<td>75</td>
<td>80 ± 0.2</td>
<td>[Colvin &amp; Hilton, 1981; Ahmed &amp; Hombal, 1984; Moore, 1991]</td>
</tr>
<tr>
<td>cytarabine</td>
<td>tra</td>
<td>20</td>
<td>20</td>
<td>29 ± 0.9</td>
<td>[Capizzi et al., 1991]</td>
</tr>
<tr>
<td>dexamethasone</td>
<td>tra</td>
<td>80-90</td>
<td>90</td>
<td>79 ± 0.2</td>
<td>[Brophy et al., 1983]</td>
</tr>
<tr>
<td>diazepam</td>
<td>tra</td>
<td>85-100</td>
<td>93</td>
<td>76 ± 0.1</td>
<td>[Reidenberg et al., 1978]</td>
</tr>
<tr>
<td>dicloxacillin</td>
<td>tra</td>
<td>35-76</td>
<td>56</td>
<td>75 ± 0.7</td>
<td>[Donowitz &amp; Mandell, 1988]</td>
</tr>
<tr>
<td>didanosine</td>
<td>tra</td>
<td>40-50</td>
<td>45</td>
<td>74 ± 0.1</td>
<td>[Faulds &amp; Brogden, 1992; Perry &amp; Balfour, 1996]</td>
</tr>
<tr>
<td>disopyramide</td>
<td>tra</td>
<td>72-94</td>
<td>83</td>
<td>70 ± 0.1</td>
<td>[Cook et al., 1993]</td>
</tr>
<tr>
<td>dolasetron</td>
<td>val</td>
<td>90</td>
<td>90</td>
<td>62 ± 0.3</td>
<td>[Dempsey et al., 1996]</td>
</tr>
<tr>
<td>domperidone</td>
<td>tra</td>
<td>15-17</td>
<td>16</td>
<td>54 ± 0.5</td>
<td>[Heykants et al., 1981]</td>
</tr>
<tr>
<td>doxapram</td>
<td>tra</td>
<td>60</td>
<td>60</td>
<td>46 ± 0.2</td>
<td>[Barrington et al., 1987]</td>
</tr>
<tr>
<td>doxepin</td>
<td>tes</td>
<td>30</td>
<td>30</td>
<td>53 ± 0.2</td>
<td>[Abernethy &amp; Todd, 1986]</td>
</tr>
<tr>
<td>doxorubicin</td>
<td>tes</td>
<td>&lt;5</td>
<td>5</td>
<td>29 ± 0.1</td>
<td>[Pourtier-Manzanedo et al., 1995]</td>
</tr>
<tr>
<td>doxycycline</td>
<td>tra</td>
<td>93</td>
<td>93</td>
<td>65 ± 0.2</td>
<td>[Saivin &amp; Houin, 1988]</td>
</tr>
<tr>
<td>Drug</td>
<td>Use</td>
<td>Range (%)</td>
<td>Target (%)</td>
<td>Predicted ± SD (%)</td>
<td>Reference</td>
</tr>
<tr>
<td>------------------</td>
<td>-------</td>
<td>-----------------</td>
<td>------------</td>
<td>--------------------</td>
<td>---------------------------------------------------------------------------</td>
</tr>
<tr>
<td>enalapril</td>
<td>tra</td>
<td>50-70</td>
<td>60</td>
<td>51 ± 0.7</td>
<td>[Gomez et al., 1985; Todd &amp; Heel, 1986; Todd &amp; Goa, 1992]</td>
</tr>
<tr>
<td>ethambutol</td>
<td>tra</td>
<td>69-85</td>
<td>77</td>
<td>66 ± 0.7</td>
<td>[Strauss &amp; Erhardt, 1970]</td>
</tr>
<tr>
<td>ethinyloestradiol</td>
<td>tra</td>
<td>40-50</td>
<td>45</td>
<td>24 ± 0.0</td>
<td>[Goldzieher &amp; Brody, 1990]</td>
</tr>
<tr>
<td>etoposide</td>
<td>tra</td>
<td>45-50</td>
<td>48</td>
<td>23 ± 0.0</td>
<td>[Toffoli et al., 2001]</td>
</tr>
<tr>
<td>etidronate</td>
<td>tra</td>
<td>5</td>
<td>5</td>
<td>17 ± 0.2</td>
<td>[Gural et al., 1985]</td>
</tr>
<tr>
<td>famotidine</td>
<td>tra</td>
<td>40-50</td>
<td>45</td>
<td>53 ± 0.6</td>
<td>[James &amp; Kearns, 1996]</td>
</tr>
<tr>
<td>felodipine</td>
<td>tes</td>
<td>20</td>
<td>20</td>
<td>23 ± 0.0</td>
<td>[Todd &amp; Faulds, 1992]</td>
</tr>
<tr>
<td>fexofenadine</td>
<td>tra</td>
<td>30</td>
<td>30</td>
<td>30 ± 1.1</td>
<td>[Simons et al., 1996]</td>
</tr>
<tr>
<td>finasteride</td>
<td>tra</td>
<td>80</td>
<td>80</td>
<td>62 ± 0.1</td>
<td>[Perry &amp; Garson Jr, 1992]</td>
</tr>
<tr>
<td>flecainide</td>
<td>tra</td>
<td>85-90</td>
<td>88</td>
<td>64 ± 0.3</td>
<td>[Winkelmann &amp; Leinberger, 1987]</td>
</tr>
<tr>
<td>fluconazole</td>
<td>tra</td>
<td>&gt;90</td>
<td>90</td>
<td>80 ± 0.2</td>
<td>[Grant &amp; Clissold, 1990]</td>
</tr>
<tr>
<td>flucytosine</td>
<td>tra</td>
<td>75-89</td>
<td>82</td>
<td>71 ± 0.3</td>
<td>[Lyman &amp; Walsh, 1992]</td>
</tr>
<tr>
<td>flunitrazepam</td>
<td>tra</td>
<td>90</td>
<td>90</td>
<td>102 ± 0.4</td>
<td>[Boxenbaum et al., 1978]</td>
</tr>
<tr>
<td>fluorouracil</td>
<td>tra</td>
<td>30</td>
<td>30</td>
<td>77 ± 0.2</td>
<td>[Diasio &amp; Harris, 1989]</td>
</tr>
<tr>
<td>fluvasatin</td>
<td>tra</td>
<td>24-24</td>
<td>24</td>
<td>51 ± 0.2</td>
<td>[Jokubaitis, 1996]</td>
</tr>
<tr>
<td>furosemide</td>
<td>tra</td>
<td>60-67</td>
<td>64</td>
<td>82 ± 0.5</td>
<td>[Cutler et al., 1974; Ponto &amp; Schoenwald, 1990a; Ponto &amp; Schoenwald, 1990b]</td>
</tr>
<tr>
<td>gabapentin</td>
<td>tra</td>
<td>50-70</td>
<td>60</td>
<td>72 ± 0.2</td>
<td>[Goa &amp; Sorkin, 1993]</td>
</tr>
<tr>
<td>gemfibrozil</td>
<td>tra</td>
<td>90</td>
<td>90</td>
<td>62 ± 0.3</td>
<td>[Todd &amp; Ward, 1988]</td>
</tr>
<tr>
<td>glibenclamide</td>
<td>tra</td>
<td>80</td>
<td>80</td>
<td>47 ± 0.5</td>
<td>[Pearson, 1985; Schwinghammer et al., 1991]</td>
</tr>
<tr>
<td>glipizide</td>
<td>tra</td>
<td>90</td>
<td>90</td>
<td>58 ± 0.9</td>
<td>[Wahlin-Boll et al., 1982; Kradjan et al., 1989; Kradjan et al., 1995]</td>
</tr>
<tr>
<td>haloperidol</td>
<td>tra</td>
<td>70</td>
<td>70</td>
<td>76 ± 0.2</td>
<td>[Morselli et al., 1982]</td>
</tr>
<tr>
<td>hydralazine</td>
<td>tes</td>
<td>30-50</td>
<td>40</td>
<td>57 ± 0.5</td>
<td>[Sproat &amp; Lopez, 1995]</td>
</tr>
<tr>
<td>hydrochlorthiazide</td>
<td>val</td>
<td>60-80</td>
<td>70</td>
<td>44 ± 0.6</td>
<td>[Allen et al., 1982]</td>
</tr>
<tr>
<td>ibuprofen</td>
<td>tra</td>
<td>85-100</td>
<td>93</td>
<td>76 ± 0.3</td>
<td>[Davies, 1998]</td>
</tr>
<tr>
<td>idarubicin</td>
<td>val</td>
<td>20-30</td>
<td>25</td>
<td>44 ± 0.5</td>
<td>[Hollingshead &amp; Faulds, 1991; Borchmann et al., 1997]</td>
</tr>
<tr>
<td>imipramine</td>
<td>tra</td>
<td>26-68</td>
<td>47</td>
<td>40 ± 0.1</td>
<td>[Nagy &amp; Johansson, 1975]</td>
</tr>
<tr>
<td>indapamide</td>
<td>tes</td>
<td>90-100</td>
<td>90</td>
<td>95 ± 0.3</td>
<td>[MIMS Australia, 1999]</td>
</tr>
<tr>
<td>Drug</td>
<td>Use a</td>
<td>Range b (%)</td>
<td>Target c (%)</td>
<td>Predicted d ± SD (%)</td>
<td>Reference</td>
</tr>
<tr>
<td>-----------------------</td>
<td>-------</td>
<td>-------------</td>
<td>--------------</td>
<td>----------------------</td>
<td>-----------</td>
</tr>
<tr>
<td>indomethacin</td>
<td>tra</td>
<td>98</td>
<td>98</td>
<td>80 ± 0.4</td>
<td>[Brooks et al., 1975; Groth &amp; Dunn, 1986]</td>
</tr>
<tr>
<td>irbesartan</td>
<td>tra</td>
<td>60-80</td>
<td>70</td>
<td>61 ± 0.8</td>
<td>[Munger &amp; Furniss, 1996]</td>
</tr>
<tr>
<td>isosorbide dinitrate</td>
<td>tra</td>
<td>25</td>
<td>25</td>
<td>28 ± 0.7</td>
<td>[Parker et al., 1987; Nakashima et al., 1990]</td>
</tr>
<tr>
<td>isosorbide mononitrate</td>
<td>tra</td>
<td>80-100</td>
<td>90</td>
<td>58 ± 0.1</td>
<td>[Abshagen, 1992]</td>
</tr>
<tr>
<td>ketamine</td>
<td>tra</td>
<td>20</td>
<td>20</td>
<td>53 ± 0.7</td>
<td>[Clements &amp; Nimmo, 1981]</td>
</tr>
<tr>
<td>ketoprofen</td>
<td>tra</td>
<td>90-100</td>
<td>95</td>
<td>89 ± 0.4</td>
<td>[Cooper, 1988]</td>
</tr>
<tr>
<td>labetalol</td>
<td>val</td>
<td>30</td>
<td>30</td>
<td>71 ± 0.6</td>
<td>[MacCarthy &amp; Bloomfield, 1983; McNeil &amp; Louis, 1984]</td>
</tr>
<tr>
<td>lamivudine</td>
<td>tra</td>
<td>66-87</td>
<td>77</td>
<td>61 ± 0.3</td>
<td>[Perry &amp; Faulds, 1997; Johnson et al., 1998]</td>
</tr>
<tr>
<td>lamotrigine</td>
<td>val</td>
<td>95</td>
<td>95</td>
<td>77 ± 0.1</td>
<td>[Cohen et al., 1987; Fitton &amp; Goa, 1995]</td>
</tr>
<tr>
<td>lansoprazole</td>
<td>tra</td>
<td>80-90</td>
<td>85</td>
<td>80 ± 0.3</td>
<td>[Chun et al., 1995; Doan et al., 2001]</td>
</tr>
<tr>
<td>levodopa</td>
<td>tra</td>
<td>44</td>
<td>44</td>
<td>61 ± 0.4</td>
<td>[Nutt &amp; Fellman, 1984]</td>
</tr>
<tr>
<td>lidocaine</td>
<td>tra</td>
<td>24-46</td>
<td>35</td>
<td>61 ± 0.5</td>
<td>[Wing et al., 1984]</td>
</tr>
<tr>
<td>lisinopril</td>
<td>tes</td>
<td>40</td>
<td>40</td>
<td>55 ± 0.3</td>
<td>[Gomez et al., 1987]</td>
</tr>
<tr>
<td>lithium carbonate</td>
<td>tra</td>
<td>100</td>
<td>100</td>
<td>84 ± 0.6</td>
<td>[Belanger et al., 1992]</td>
</tr>
<tr>
<td>loperamide</td>
<td>tra</td>
<td>&lt;40</td>
<td>40</td>
<td>35 ± 0.4</td>
<td>[MIMS Australia, 1999]</td>
</tr>
<tr>
<td>lorazepam</td>
<td>tes</td>
<td>95</td>
<td>95</td>
<td>94 ± 0.4</td>
<td>[Greenblatt et al., 1979]</td>
</tr>
<tr>
<td>losartan</td>
<td>val</td>
<td>33</td>
<td>33</td>
<td>40 ± 0.4</td>
<td>[Brunner et al., 1992]</td>
</tr>
<tr>
<td>meperidine</td>
<td>tes</td>
<td>50-60</td>
<td>55</td>
<td>52 ± 0.7</td>
<td>[Miller &amp; Jick, 1978]</td>
</tr>
<tr>
<td>mercaptopurine</td>
<td>tra</td>
<td>16-46</td>
<td>31</td>
<td>42 ± 0.6</td>
<td>[Elion, 1967; Loo et al., 1968; Bostrom &amp; Erdmann, 1993]</td>
</tr>
<tr>
<td>metformin</td>
<td>tra</td>
<td>50-60</td>
<td>55</td>
<td>70 ± 0.4</td>
<td>[Dunn &amp; Peters, 1995]</td>
</tr>
<tr>
<td>methotrexate</td>
<td>tra</td>
<td>60-70</td>
<td>68</td>
<td>48 ± 0.1</td>
<td>[Bleyer, 1977]</td>
</tr>
<tr>
<td>methylprednisolone</td>
<td>tes</td>
<td>79-82</td>
<td>81</td>
<td>71 ± 0.5</td>
<td>[Tormatore et al., 1994]</td>
</tr>
<tr>
<td>metoclopramide</td>
<td>val</td>
<td>30-90</td>
<td>60</td>
<td>70 ± 0.3</td>
<td>[Lauritsen et al., 1990]</td>
</tr>
<tr>
<td>metoprolol</td>
<td>tra</td>
<td>50-50</td>
<td>50</td>
<td>59 ± 0.1</td>
<td>[Shore et al., 1981]</td>
</tr>
<tr>
<td>metronidazole</td>
<td>tra</td>
<td>90</td>
<td>90</td>
<td>82 ± 0.8</td>
<td>[Freeman et al., 1997]</td>
</tr>
<tr>
<td>mexiletine</td>
<td>tra</td>
<td>80-100</td>
<td>90</td>
<td>70 ± 0.4</td>
<td>[Monk &amp; Brogden, 1990]</td>
</tr>
<tr>
<td>mianserin</td>
<td>tra</td>
<td>30</td>
<td>30</td>
<td>32 ± 0.1</td>
<td>[Timmer et al., 1985]</td>
</tr>
<tr>
<td>minocycline</td>
<td>tra</td>
<td>100</td>
<td>100</td>
<td>74 ± 0.3</td>
<td>[Meyer, 1996]</td>
</tr>
<tr>
<td>misoprostol</td>
<td>tra</td>
<td>80</td>
<td>80</td>
<td>69 ± 0.2</td>
<td>[Walt, 1992]</td>
</tr>
<tr>
<td>morphine</td>
<td>tra</td>
<td>20-33</td>
<td>27</td>
<td>28 ± 0.2</td>
<td>[Brunk &amp; Delle, 1974; June et al., 1995; Olkkola et al., 1995; Beyssac et al., 1998]</td>
</tr>
<tr>
<td>Drug</td>
<td>Use</td>
<td>Range (%)</td>
<td>Target (%)</td>
<td>Predicted (± SD) (%)</td>
<td>Reference</td>
</tr>
<tr>
<td>------------</td>
<td>------</td>
<td>-----------</td>
<td>------------</td>
<td>----------------------</td>
<td>--------------------------------------------------------------------------</td>
</tr>
<tr>
<td>naloxone</td>
<td>val</td>
<td>2-10</td>
<td>6</td>
<td>42 ± 0.2</td>
<td>[Anonymous, 1990]</td>
</tr>
<tr>
<td>naltrexone</td>
<td>val</td>
<td>20-22</td>
<td>20</td>
<td>38 ± 0.7</td>
<td>[Hussain et al., 1987]</td>
</tr>
<tr>
<td>nifdefipine</td>
<td>val</td>
<td>45-85</td>
<td>65</td>
<td>62 ± 0.1</td>
<td>[Ferner et al., 1990]</td>
</tr>
<tr>
<td>nimodipine</td>
<td>tra</td>
<td>13</td>
<td>13</td>
<td>59 ± 0.2</td>
<td>[Ramoska et al., 1990]</td>
</tr>
<tr>
<td>nitrazepam</td>
<td>tra</td>
<td>80</td>
<td>80</td>
<td>101 ± 0.5</td>
<td>[Ochs et al., 1983]</td>
</tr>
<tr>
<td>nizatidine</td>
<td>tra</td>
<td>90</td>
<td>90</td>
<td>71 ± 0.2</td>
<td>[Callaghan et al., 1987; Vargas et al., 1988]</td>
</tr>
<tr>
<td>naloxone</td>
<td>tra</td>
<td>45</td>
<td>45</td>
<td>79 ± 0.1</td>
<td>[Nilsson-Ehle &amp; Ljungberg, 1991]</td>
</tr>
<tr>
<td>nortriptyline</td>
<td>tra</td>
<td>46-56</td>
<td>51</td>
<td>44 ± 0.0</td>
<td>[Dawling et al., 1980; Lipper &amp; Gaynor, 1995; Kvist et al., 2001]</td>
</tr>
<tr>
<td>omeprazole</td>
<td>tra</td>
<td>40-60</td>
<td>50</td>
<td>71 ± 0.8</td>
<td>[Andersson, 1991; Massoomi et al., 1993]</td>
</tr>
<tr>
<td>ondansetron</td>
<td>tra</td>
<td>48-75</td>
<td>62</td>
<td>60 ± 0.7</td>
<td>[Chaffee &amp; Tankanow, 1991; Roila &amp; Del Favero, 1995; Simpson &amp; Hicks, 1996; Jann et al., 1998]</td>
</tr>
<tr>
<td>oriprenaline</td>
<td>tra</td>
<td>40</td>
<td>40</td>
<td>76 ± 0.5</td>
<td>[MIMS Australia, 1999]</td>
</tr>
<tr>
<td>oxprenolol</td>
<td>tra</td>
<td>50</td>
<td>50</td>
<td>58 ± 0.1</td>
<td>[Leucuta et al., 1998]</td>
</tr>
<tr>
<td>oxycodone</td>
<td>tra</td>
<td>50</td>
<td>50</td>
<td>31 ± 0.1</td>
<td>[Kalso &amp; Vainio, 1990]</td>
</tr>
<tr>
<td>pamidronate</td>
<td>tra</td>
<td>1</td>
<td>1</td>
<td>-3 ± 0.1</td>
<td>[Fitton &amp; McTavish, 1991]</td>
</tr>
<tr>
<td>pantoprazole</td>
<td>tes</td>
<td>80-97</td>
<td>89</td>
<td>75 ± 0.2</td>
<td>[Huber et al., 1996]</td>
</tr>
<tr>
<td>paracetamol</td>
<td>val</td>
<td>58-68</td>
<td>63</td>
<td>77 ± 0.2</td>
<td>[Muir et al., 1997a]</td>
</tr>
<tr>
<td>pethidine</td>
<td>tra</td>
<td>60</td>
<td>60</td>
<td>57 ± 0.6</td>
<td>[Shih et al., 1994]</td>
</tr>
<tr>
<td>phenobarbital</td>
<td>tra</td>
<td>70-90</td>
<td>80</td>
<td>69 ± 0.3</td>
<td>[MIMS Australia, 1999]</td>
</tr>
<tr>
<td>phenytoin</td>
<td>tra</td>
<td>98</td>
<td>98</td>
<td>74 ± 0.1</td>
<td>[Gugler et al., 1992]</td>
</tr>
<tr>
<td>pindolol</td>
<td>tra</td>
<td>90</td>
<td>90</td>
<td>63 ± 0.5</td>
<td>[Chau et al., 1977]</td>
</tr>
<tr>
<td>pravastatin</td>
<td>tra</td>
<td>17-34</td>
<td>26</td>
<td>67 ± 0.6</td>
<td>[McTavish &amp; Sorkin, 1991]</td>
</tr>
<tr>
<td>prazosin</td>
<td>tra</td>
<td>43-82</td>
<td>63</td>
<td>67 ± 0.3</td>
<td>[Reece, 1980; Rostock et al., 1985; Mikami &amp; Ogihara, 1990]</td>
</tr>
<tr>
<td>primidone</td>
<td>tra</td>
<td>60-80</td>
<td>70</td>
<td>68 ± 0.2</td>
<td>[Booker et al., 1970; Watanabe et al., 1977]</td>
</tr>
<tr>
<td>procainamide</td>
<td>tra</td>
<td>67-99</td>
<td>83</td>
<td>78 ± 0.1</td>
<td>[Koch-Weser, 1977]</td>
</tr>
<tr>
<td>prochlorperazine</td>
<td>tra</td>
<td>20</td>
<td>20</td>
<td>44 ± 0.7</td>
<td>[Tokola, 1988]</td>
</tr>
<tr>
<td>promethazine</td>
<td>tes</td>
<td>25</td>
<td>25</td>
<td>38 ± 0.2</td>
<td>[Schwinghammer et al., 1984]</td>
</tr>
<tr>
<td>propranolol</td>
<td>tra</td>
<td>26-46</td>
<td>36</td>
<td>68 ± 0.1</td>
<td>[Barnwell et al., 1996]</td>
</tr>
<tr>
<td>propylthiouracil</td>
<td>val</td>
<td>80-90</td>
<td>85</td>
<td>88 ± 0.3</td>
<td>[Duarte et al., 2000]</td>
</tr>
<tr>
<td>quinapril</td>
<td>tra</td>
<td>50-60</td>
<td>55</td>
<td>39 ± 0.0</td>
<td>[Horvath et al., 1990]</td>
</tr>
<tr>
<td>quinidine</td>
<td>tes</td>
<td>54-88</td>
<td>70</td>
<td>67 ± 0.2</td>
<td>[Verme et al., 1992]</td>
</tr>
<tr>
<td>ramipril</td>
<td>tra</td>
<td>50-60</td>
<td>55</td>
<td>57 ± 0.3</td>
<td>[Kindler et al., 1989]</td>
</tr>
<tr>
<td>ranitidine</td>
<td>val</td>
<td>40-80</td>
<td>60</td>
<td>75 ± 0.2</td>
<td>[Grant et al., 1989]</td>
</tr>
<tr>
<td>Drug</td>
<td>Use</td>
<td>Range (b)</td>
<td>Target (c)</td>
<td>Predicted (d)</td>
<td>Reference</td>
</tr>
<tr>
<td>--------------------</td>
<td>-----</td>
<td>-----------</td>
<td>------------</td>
<td>---------------</td>
<td>---------------------------------------------------------------------------</td>
</tr>
<tr>
<td>ribavirin</td>
<td>tra</td>
<td>40-64</td>
<td>52</td>
<td>45 ± 0.5</td>
<td>[Smee et al., 1981]</td>
</tr>
<tr>
<td>salicylic acid</td>
<td>tra</td>
<td>100</td>
<td>100</td>
<td>79 ± 0.2</td>
<td>[Gatti et al., 1989]</td>
</tr>
<tr>
<td>selegiline</td>
<td>tra</td>
<td>20</td>
<td>20</td>
<td>0 ± 0.2</td>
<td>[Trebin &amp; Gordon, 1995]</td>
</tr>
<tr>
<td>simvastatin</td>
<td>tra</td>
<td>5</td>
<td>5</td>
<td>8 ± 0.4</td>
<td>[Grau et al., 1996]</td>
</tr>
<tr>
<td>sotalol</td>
<td>tra</td>
<td>95</td>
<td>95</td>
<td>81 ± 0.4</td>
<td>[Kahela et al., 1979; Fitton &amp; Sorkin, 1993]</td>
</tr>
<tr>
<td>spironolactone</td>
<td>tra</td>
<td>70-90</td>
<td>80</td>
<td>87 ± 0.1</td>
<td>[Bartle et al., 1979]</td>
</tr>
<tr>
<td>sulfamethoxazole</td>
<td>tra</td>
<td>80-90</td>
<td>85</td>
<td>69 ± 0.4</td>
<td>[Lewin et al., 1973]</td>
</tr>
<tr>
<td>sulfisoxazole</td>
<td>tra</td>
<td>100</td>
<td>100</td>
<td>73 ± 0.4</td>
<td>[Garrett et al., 1981]</td>
</tr>
<tr>
<td>sumatriptan</td>
<td>tra</td>
<td>15-20</td>
<td>18</td>
<td>71 ± 0.7</td>
<td>[Warner et al., 1995; Cosson &amp; Fuseau, 1999]</td>
</tr>
<tr>
<td>tacrine</td>
<td>tra</td>
<td>30-40</td>
<td>35</td>
<td>51 ± 0.4</td>
<td>[Krall et al., 1999]</td>
</tr>
<tr>
<td>temazepam</td>
<td>tra</td>
<td>90</td>
<td>90</td>
<td>91 ± 0.3</td>
<td>[Kroboth et al., 1985; Hosie &amp; Nimmo, 1991]</td>
</tr>
<tr>
<td>terbutaline</td>
<td>tra</td>
<td>15</td>
<td>15</td>
<td>46 ± 0.7</td>
<td>[Borgström et al., 1989]</td>
</tr>
<tr>
<td>testosteren</td>
<td>tra</td>
<td>5</td>
<td>5</td>
<td>45 ± 0.2</td>
<td>[Täuber et al., 1986]</td>
</tr>
<tr>
<td>tetracycline</td>
<td>tes</td>
<td>77</td>
<td>77</td>
<td>72 ± 0.7</td>
<td>[Kramer et al., 1978; Albert et al., 1979]</td>
</tr>
<tr>
<td>theophyline</td>
<td>tra</td>
<td>80-100</td>
<td>90</td>
<td>76 ± 0.8</td>
<td>[Rojanaasthien et al., 2001]</td>
</tr>
<tr>
<td>thioprofenic acid</td>
<td>tra</td>
<td>100</td>
<td>100</td>
<td>81 ± 0.1</td>
<td>[Massa et al., 1986]</td>
</tr>
<tr>
<td>timolol</td>
<td>tra</td>
<td>30-50</td>
<td>90</td>
<td>79 ± 0.1</td>
<td>[Sutinen et al., 2000]</td>
</tr>
<tr>
<td>tolbutamide</td>
<td>tes</td>
<td>93</td>
<td>93</td>
<td>69 ± 0.5</td>
<td>[Antal et al., 1982]</td>
</tr>
<tr>
<td>triamterene</td>
<td>tra</td>
<td>50</td>
<td>50</td>
<td>55 ± 0.1</td>
<td>[Williams et al., 1987; Heimsoth et al., 1988; Mühlberg et al., 2001]</td>
</tr>
<tr>
<td>trimethoprim</td>
<td>tra</td>
<td>90-97</td>
<td>94</td>
<td>61 ± 0.2</td>
<td>[Meshi &amp; Sato, 1972; Klinker et al., 1998]</td>
</tr>
<tr>
<td>trimipramine</td>
<td>tra</td>
<td>40</td>
<td>40</td>
<td>39 ± 0.0</td>
<td>[Caillé et al., 1980]</td>
</tr>
<tr>
<td>tropisetron</td>
<td>tra</td>
<td>60</td>
<td>60</td>
<td>59 ± 0.2</td>
<td>[Kees et al., 2001]</td>
</tr>
<tr>
<td>valproic acid</td>
<td>tra</td>
<td>95-100</td>
<td>100</td>
<td>87 ± 0.7</td>
<td>[Davis et al., 1994]</td>
</tr>
<tr>
<td>verapamil</td>
<td>tra</td>
<td>20-35</td>
<td>28</td>
<td>39 ± 0.3</td>
<td>[McTavish, 1989]</td>
</tr>
<tr>
<td>warfarin</td>
<td>tra</td>
<td>98</td>
<td>98</td>
<td>94 ± 0.3</td>
<td>[Müller et al., 1988]</td>
</tr>
<tr>
<td>zalcitabine</td>
<td>tra</td>
<td>70-88</td>
<td>79</td>
<td>58 ± 0.1</td>
<td>[Adams et al., 1998]</td>
</tr>
<tr>
<td>zidovudine</td>
<td>tra</td>
<td>60-65</td>
<td>63</td>
<td>49 ± 0.4</td>
<td>[Guarino et al., 1998]</td>
</tr>
</tbody>
</table>

*a* data subset for ANN model: tra – training, val – validation, tes – testing; *b* range of bioavailability values from literature; *c* target bioavailability values used for ANN models; *d* bioavailability predicted by ANN.
6.2.3 Input Variable Selection

Both manual and automatic pruning techniques were used to reduce the number of input variables. Initially radial-basis function models with different topology were trained and tested. Best models were then selected for sensitivity analyses and examination of the input neuron activation levels. Sensitivity analysis of inputs was used to identify significance of individual molecular descriptors and to select descriptors that were considered the most important (see Section 2.4). Based on results of sensitivity analyses, inputs with sensitivity ratios less than one were eliminated sequentially from the model. As the number of input variables was reduced, descriptors with sensitivity greater than one but low relative values were removed manually. Activations can be either positive or negative and represent the strength of the output from a given neuron. High absolute activations indicate a substantial contribution to the model and low absolute activations indicate the opposite. Input neurons displaying a zero activation do not contribute to the system at all and these were manually pruned from the model. The ANN program also utilised regularisation and search algorithms for automatic descriptor selection.

6.2.4 Network Construction

Statistica Neural Networks 4.0 [StatSoft Inc, 2000] was used to construct a three layered radial-basis function ANN containing a bias neuron in each layer and a single neuron in the output layer. The calculated molecular descriptors were used as inputs to the ANN and the target output was the bioavailability data. Network weights and biases were initialised with random values before each training run, and all runs were performed in replicates of five. The training set was used to train
the ANN, the validation set to evaluate ANN performance and monitor overtraining, and the testing set to evaluate the predictive ability of the trained model. Training was stopped when the training root mean squared (RMS) error failed to improve over a given number of training cycles and when the testing RMS error started to increase. The training and test correlation coefficients ($r_t$ and $r_{val}$ respectively) were used to evaluate the overall quality of a particular subset of descriptors and the corresponding network topology.

All 76 descriptors were included in the initial model after which pruning was implemented. Repeated training runs for each configuration were necessary to avoid undertraining and also to prevent the network from falling into local minima. If the network was undertrained and did not achieve an acceptable performance level, the model was discarded and another replicate was performed.

### 6.2.4.1 Modified Efficiency Ratio

Predictive network performance was evaluated according to an efficiency ratio, ER (Section 4.2.5). In the present study ER was defined as the ratio of the testing correlation coefficient ($r_{tes}$) to the training correlation ($r_t$) coefficient (Equation 6-31):

$$\text{ER} = \frac{r_{tes}}{r_t} \quad \text{Equation 6-31}$$

This modification of Equation 4-23 utilised $r_{tes}$ instead of $r_{cv}$ since the predictive ability of the present model was determined by the testing rather than validation data.
6.2.5 Stepwise Regression Modeling

In order to compare multilinear regression and ANN approaches a forward stepwise regression (SWR) model was constructed. The forward stepwise method employs a combination of the forward entry of independent variables and backward removal of insignificant variables. The best single predictor, which is the most significant variable, was used for the initial linear regression step. Next, descriptors were added one at a time, always adding the one that most improved the fit, until the fit was not significantly improved. Once all the significant variables were determined the regression equation was assembled.

6.2.5.1 Regression Data

The same bioavailability data was used for the stepwise regression model as the ANN model. However, the testing subset consisted of 10 different compounds: fexofenadine, flecainide, fluvastatin, flucytosine, glipizide, haloperidol, hyoscine, imipramine, indomethacin, and trimethoprim. The remaining 159 compounds were used for model training.

6.3 Results and Discussion – ANN Model

6.3.1 Descriptor Pruning

Initially, a neural network consisting of 76 input variables, one hidden layer, and one output neuron for the target bioavailability was used. Following pruning, the number of inputs was reduced from 76 to 66, 47, 27, 24, 19, 12, 10, 9, 7 and finally to 4 inputs. ER increased as descriptors were removed and peaked at the 47...
input model after which it began to decrease (Table 6-18). A second peak was seen for the 10 input model and further pruning caused a decrease in ER.

The ANN model with 10 input descriptors and 31 hidden neurons was found to have the highest $r_t$ value and a high ER. Other architectures containing different numbers of hidden neurons were examined but they produced poorer quality ANN models and worse predictions. Since the 10 input model with achieved relatively high $r_{les}$ and ER values it was taken as the optimum model and subjected to further analysis.

**Table 6-18.** ANN bioavailability model summary over pruning.

<table>
<thead>
<tr>
<th>Model</th>
<th>ER</th>
<th>$r_t \pm SD$</th>
<th>$r_{val} \pm SD$</th>
<th>$r_{les} \pm SD$</th>
</tr>
</thead>
<tbody>
<tr>
<td>76-29-1</td>
<td>0.52</td>
<td>0.668 ± 0.000</td>
<td>0.801 ± 0.003</td>
<td>0.349 ± 0.008</td>
</tr>
<tr>
<td>66-18-1</td>
<td>0.58</td>
<td>0.629 ± 0.002</td>
<td>0.782 ± 0.001</td>
<td>0.365 ± 0.020</td>
</tr>
<tr>
<td>47-20-1</td>
<td>1.11</td>
<td>0.608 ± 0.004</td>
<td>0.771 ± 0.004</td>
<td>0.678 ± 0.009</td>
</tr>
<tr>
<td>27-28-1</td>
<td>1.00</td>
<td>0.688 ± 0.001</td>
<td>0.845 ± 0.002</td>
<td>0.686 ± 0.005</td>
</tr>
<tr>
<td>24-31-1</td>
<td>0.88</td>
<td>0.702 ± 0.002</td>
<td>0.850 ± 0.001</td>
<td>0.620 ± 0.007</td>
</tr>
<tr>
<td>19-44-1</td>
<td>0.71</td>
<td>0.721 ± 0.004</td>
<td>0.865 ± 0.010</td>
<td>0.509 ± 0.021</td>
</tr>
<tr>
<td>12-54-1</td>
<td>0.45</td>
<td>0.779 ± 0.005</td>
<td>0.887 ± 0.000</td>
<td>0.354 ± 0.003</td>
</tr>
<tr>
<td>10-31-1</td>
<td>0.92</td>
<td>0.736 ± 0.000</td>
<td>0.897 ± 0.004</td>
<td>0.680 ± 0.002</td>
</tr>
<tr>
<td>9-44-1</td>
<td>0.73</td>
<td>0.760 ± 0.002</td>
<td>0.834 ± 0.002</td>
<td>0.552 ± 0.004</td>
</tr>
<tr>
<td>7-37-1</td>
<td>0.70</td>
<td>0.732 ± 0.000</td>
<td>0.870 ± 0.007</td>
<td>0.513 ± 0.002</td>
</tr>
<tr>
<td>4-8-1</td>
<td>0.76</td>
<td>0.469 ± 0.004</td>
<td>0.547 ± 0.011</td>
<td>0.354 ± 0.017</td>
</tr>
</tbody>
</table>

$r_t$ – training correlation; $r_{val}$ – validation correlation; $r_{les}$ – testing correlation.

### 6.3.2 Statistical Analysis

Since the variance of experimental bioavailability values between subsets was not significant ($p>0.7$) and normality for all subsets was able to be assumed ($p>0.10$), a high ANOVA statistic indicated that compounds selected in the testing and validation subsets were representative of the training set (Table 6-19). Statistical analysis of the descriptor data was not performed prior to pruning since removal
of a large number of descriptors would greatly change the apparent descriptor
space. However, analysis of the optimum descriptor set was performed and
revealed low differences in variance between subsets ($0.15 < p < 0.80$) and mostly
normal distributions. Appropriate parametric and nonparametric analyses of
individual descriptors did not reveal any significant differences between
information contained in training, validation or testing subsets ($0.15 < p < 0.97$). A
Bonferroni post-hoc analysis revealed similar results (Appendix A2).
<table>
<thead>
<tr>
<th>Data</th>
<th>Training normality&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Validation normality&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Testing normality&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Variance&lt;sup&gt;c&lt;/sup&gt;</th>
<th>ANOVA&lt;sup&gt;d&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bioavailability</td>
<td>0.11</td>
<td>0.12</td>
<td>0.32</td>
<td>0.74</td>
<td>0.91</td>
</tr>
<tr>
<td>Difference index &lt;sup&gt;2&lt;/sup&gt;Δ</td>
<td>0.03</td>
<td>0.06</td>
<td>0.23</td>
<td>0.66</td>
<td>0.71</td>
</tr>
<tr>
<td>Kappa shape index &lt;sup&gt;2&lt;/sup&gt;κ</td>
<td>0.1</td>
<td>0.35</td>
<td>0.03</td>
<td>0.48</td>
<td>0.83</td>
</tr>
<tr>
<td>Aromatic ring counts</td>
<td>&gt;0.01</td>
<td>0.05</td>
<td>0.01</td>
<td>0.7</td>
<td>0.62&lt;sup&gt;e&lt;/sup&gt;</td>
</tr>
<tr>
<td>Molar refractivity</td>
<td>0.2</td>
<td>0.98</td>
<td>0.31</td>
<td>0.22</td>
<td>0.5</td>
</tr>
<tr>
<td>Dielectric energy</td>
<td>&gt;0.01</td>
<td>0.01</td>
<td>0.04</td>
<td>0.18</td>
<td>0.17&lt;sup&gt;e&lt;/sup&gt;</td>
</tr>
<tr>
<td>Electron affinity</td>
<td>&gt;0.01</td>
<td>0.08</td>
<td>0.4</td>
<td>0.76</td>
<td>0.57</td>
</tr>
<tr>
<td>Conformation minimum energy</td>
<td>0.2</td>
<td>0.66</td>
<td>0.94</td>
<td>0.8</td>
<td>0.97</td>
</tr>
<tr>
<td>Polar surface area</td>
<td>0.2</td>
<td>0.24</td>
<td>0.36</td>
<td>0.16</td>
<td>0.71</td>
</tr>
<tr>
<td>Connectivity index &lt;sup&gt;1&lt;/sup&gt;χ</td>
<td>0.2</td>
<td>0.89</td>
<td>0.55</td>
<td>0.71</td>
<td>0.55</td>
</tr>
<tr>
<td>Valence connectivity index &lt;sup&gt;4&lt;/sup&gt;χ&lt;sup&gt;v&lt;/sup&gt;</td>
<td>0.15</td>
<td>0.44</td>
<td>0.59</td>
<td>0.75</td>
<td>0.53</td>
</tr>
</tbody>
</table>

<sup>a</sup>Kolmogorov-Smirnov test of normality; <sup>b</sup>Shapiro-Wilk test of normality; <sup>c</sup>Levene’s homogeneity of variance; <sup>d</sup>analysis of variance (ANOVA) except where indicated; <sup>e</sup>Kruskal-Wallis test.
6.3.3 Descriptor Analysis

Due to uncertainty in the weight matrices of conventional back-propagation ANNs conclusions about relationships between input and output variables are often difficult to make. Use of radial-basis function ANNs, however, does allow these relationships to be examined [Derks et al., 1995]. The 10 descriptors in the optimum model and their overall sensitivity ranks are given in Table 6-20. Sensitivity ranks indicate the relative importance of each descriptor on the final model. They were obtained by summat ing the absolute sensitivities for each descriptor over all training replicates performed. It was found that simple constitutional descriptors such as aromatic ring counts could be just as important as complex 3D descriptors such as polar surface area (PSA) and dielectric energy.

Table 6-20. Sensitivity ranks of selected descriptors in the optimum ANN bioavailability model.

<table>
<thead>
<tr>
<th>Descriptor name</th>
<th>Input Symbol</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aromatic ring counts</td>
<td>Ar.ring</td>
<td>2</td>
</tr>
<tr>
<td>Conformation minimum energy</td>
<td>Conform</td>
<td>7</td>
</tr>
<tr>
<td>Connectivity index $^1\chi$</td>
<td>Conn1</td>
<td>6</td>
</tr>
<tr>
<td>Dielectric energy</td>
<td>Dielec</td>
<td>3</td>
</tr>
<tr>
<td>Difference index $^2\Delta$</td>
<td>Diff2</td>
<td>8</td>
</tr>
<tr>
<td>Electron affinity</td>
<td>El.aff</td>
<td>5</td>
</tr>
<tr>
<td>Kappa shape index $\kappa^2$</td>
<td>Kappa2</td>
<td>10</td>
</tr>
<tr>
<td>Molar refractivity</td>
<td>MR</td>
<td>9</td>
</tr>
<tr>
<td>Polar surface area</td>
<td>PSA</td>
<td>1</td>
</tr>
<tr>
<td>Valence connectivity index $^4\chi^v$</td>
<td>Val4</td>
<td>4</td>
</tr>
</tbody>
</table>

Sensitivity ratios varied slightly between training and testing sets (Figure 6-27) although relative values and, hence, sensitivity ranks were not considerably affected. The optimum model indicated that physicochemical factors affect drug absorption and consequently bioavailability including the intrinsic solubility.
(PSA) of the drug molecule, as well as its electronic nature (electron affinity, dielectric and conformational energy, aromatic ring counts), and molecular size and shape characteristics (molar refractivity, and connectivity, difference and shape indices).

![Figure 6-27](image)

**Figure 6-27.** Training and testing sensitivities for optimum ANN bioavailability model.

Each of the descriptors in the optimum model had different effects on the predicted bioavailability (Figure 6-28). The response graphs were generated by varying the values of the descriptor in question while holding values for all the other descriptors constant. They displayed the effect of each descriptor on the predicted bioavailability.
Figure 6-28. Response graphs for optimum ANN bioavailability model descriptor set.
It has been shown that PSA with hydrogen bonding capacity (PSA and the presence of –OH groups) plays a significant role in the description of drug membrane penetration [Winiwarter et al., 1998], and that PSA can be used as a predictor of absorption [Palm, 1997]. Molecular surface area and volume are highly correlated geometrical descriptors that can provide information about contact surface, surface diffusion, absorption and information of the size of the molecules. The contact surface area can be viewed as an indicator of the extent to which the solute is exposed to intermolecular interaction with the solvent [Hermann, 1997] and has been shown to be an accurate predictor of water solubility [Bodor & Huang, 1992]. Drugs need to be in solution before they can be absorbed from the gastrointestinal tract. Therefore, as a general rule, a drug that is very poorly soluble or insoluble in water would have variable or unreliable absorption.

PSA also indicates the capacity of a compound to form hydrogen bonds. Hydrogen bonds are major forces of recognition in biochemistry and molecular pharmacology: they are an essential component of intermolecular interactions. Calculated surface characteristics of molecules have been correlated with a number of physicochemical properties of drug molecules including lipophilicity, the energy of hydration and the hydrogen bond formation capacity [Dunn III et al., 1987; Ooi et al., 1987]. An increase in the value of PSA in the optimum model corresponded to an initial positive effect on bioavailability but then caused predicted bioavailability to drop substantially (Figure 6-28i).

A dielectric material is a substance that is a poor conductor of electricity, but an efficient supporter of electrostatic fields. All molecules have surfaces, and charge
can accumulate on those surfaces. Charge accumulation can affect the formation of hydrogen bonds, which has been shown to play an important role in enzymatic catalysis [Shan & Herschlag, 1996]. The strength of hydrogen bonds between an enzyme and substrate changes over the course of a reaction. The binding energy of an enzyme is used to fix the substrate in the low-dielectric active site, from where the strength of the hydrogen bond is increased over the course of a reaction. The dielectric energy descriptor accounts for the original charge arrangement on the surface of the molecule, and thus would be indicative of the metabolic susceptibility of a drug molecule. An increase in the dielectric energy in the optimum model corresponded to an overall decrease in predicted bioavailability (Figure 6-28d). High dielectric energies resulted in negatively predicted bioavailability values. However, negative bioavailability is not possible in a physiological sense. The explanation is that the final predicted bioavailability is a combination of the influence of all the descriptors and not a product of only one descriptor. Therefore, the response graphs do not indicate the absolute value of the predicted bioavailability based on a particular descriptor but instead indicate the influence of that descriptor on bioavailability.

For a drug to be absorbed from the GI tract it must be capable of moving across cell membranes (transcellular absorption) or between the tight gaps that are formed between cells (paracellular absorption). Drug penetration through the biological membranes depends upon a number of molecular properties, such as lipophilicity, polarity, degree of ionisation and molecular size. Penetration between the cell gaps depends on molecular size and the concentration gradient. Molecular size, in general, limits the absorption of drugs through membranes.
Small, lipid insoluble substances penetrate cell membranes via the pores between aqueous phases on both sides of the membrane. The rate of such passive diffusion depends on the size of the pores, the molecular volume of the solute and the solute concentration gradient. Compounds with low molecular mass [Mennella & Beauchamp, 1991] that are not ionised and are lipophilic will have higher bioavailability simply because diffusion through pores is much easier. Calculation of molar refractivity is based upon both molar mass and density, and molar refractivity has been shown to be correlated with geometric volume [Bhattacharjee & Dasgupta, 1994]. Geometric volume relates to molecular size, indicating that molar refractivity would contribute to the observed absorption of drugs from the GI tract. The relationship between molar refractivity and bioavailability appeared to follow a cubic trend (Figure 6-28h). Hence, increasing molar refractivity caused predicted bioavailability to initially increase, then decrease, and then increase again. This demonstrated the nonlinearity apparent between the descriptor and target output spaces.

Molecular connectivity indices represent molecular structure in a manner similar to the counts of carbon atoms, but in much more general way. That is, $\chi$ indices are weighted counts of structure features with the same mathematical qualities as counts, but with much more structural information. Structural features such as size, branching, unsaturation, heteroatom content and cyclicity are encoded. The connectivity approach is fundamentally different from traditional biological QSAR methods based on assumed mechanisms and using physicochemical properties as regression variables. The connectivity method directly correlates structural information with molecular activity and not indirectly through an
Chapter 6. QSPkR for a Diverse Series

intermediate physical property. The structure base of $\chi$ indices has enabled sufficient information to be extracted from QSAR equations to allow molecules to be designed directly from those equations [Hall & Kier, 1993; Kier & Hall, 1993]. Connectivity indices up to the fourth order are known to encode various molecular properties including molecular density, branching and aromatic ring substitutions. Linear combinations of connectivity indices have been useful, especially in dealing with structurally diverse data sets [Galvez et al., 1994b]. In addition to encoding structural information, the difference index $^2\Delta$ also provides information on inductive and delocalisation effects. Even though there were slight positive gradients of the response at the extremes of the graph, there was an overall negative trend in the response of bioavailability to the $^2\Delta$ descriptor (Figure 6-28e).

The first order connectivity index, $^1\chi$ encodes single bond properties and is a weighted count of bonds, being related to the types and position of branching in the molecule. $^2\chi$ also provides information about the types and position of branching and may indicate the amount of structural flexibility of a molecule. Although it is derived from fragments of two bond lengths, $^2\chi$ is highly correlated with $^1\chi$. Structural and steric information contained in $^1\chi$ is also reflected in other descriptors related to molecular shape.

The three topological shape indices [Randic, 2001] numerically quantify molecular topology. They present information concerning the size, shape, branching pattern, cyclicity and similarity of molecular graphs. $\kappa^2$ encodes linearity of a molecule, and inclusion in the current model for bioavailability would provide structural and shape information not present in $^1\chi$. 

192
Valence connectivity indices [Estrada, 2001] use the same graph invariant as the $\chi$ indices described previously, but with modified vertex degrees to account for heteroatoms. Practical application of $\chi$ indices is heavily dependent upon the ability to deal with molecules containing heteroatoms. Valence connectivity indices are calculated using the number of valence electrons in the corresponding atom, and can be used to differentiate between heteroatoms in various functional groups. $^{4}\chi^v$ accounts for heteroatom substitution on benzene rings, the aromatic nature of which can affect solubility of a compound.

The response graphs for $^1\chi$ (Figure 6-28c), $\kappa^2$ (Figure 6-28g), and $^{4}\chi^v$ (Figure 6-28j) all display nonlinear relationships with bioavailability. The overall influences of $^1\chi$ and $^{4}\chi^v$ were similar in appearance perhaps reflecting their common origin. The more pronounced effect of $^{4}\chi^v$ was expected since $^{4}\chi^v$ is a more complex descriptor than $^1\chi$.

Although molecular solubility descriptors and topological shape indices can successfully rationalise compound solubility, they do not provide information on electronic influence through bonds or across space. Electronic properties, such as field and resonance effects, may play a role in describing the magnitude of biological activity in conjunction with structural features encoded in indexes. This can be explained by the fact that electron affinity was included in the model as a physical property that influences the chemical behaviour of the molecule. Electron affinity is the change in the total energy of a molecule when an electron is added to form a negatively charged ion. For drugs that can ionise, solubility will depend on physiological factors such as the local pH conditions within the stomach and
intestines. Resonance effects are influenced by the presence of aromatic rings. When a hydroxyl group is appended to an aromatic ring, the resultant phenol is a weak acid and is able to dissociate in water to form the corresponding phenolate anion. This dissociation is more facile due to resonance stabilisation of the phenolate, in which the negative charge delocalises into the aromatic system. Such charge delocalisation causes a decrease in the electron density of the group attached to aromatic ring and an increase in the electron density of the aromatic ring itself. Aromatic compounds are characterised by a special stability and they undergo substitution reactions more easily than addition reactions. Inclusion of aromatic ring counts in the final model would also complement the information provide by $\chi^v$. Quantum chemical descriptors further describe electronic and reactive properties of drug molecules. Minimum energy of a molecule is indicative of stability and reactivity. Increasing reactivity of a molecule corresponds to an increased potential for metabolism, which would then affect drug bioavailability.

The response graphs for aromatic ring counts (Figure 6-28a) and conformational minimum energy (Figure 6-28b) presented nonlinear relationships with bioavailability. The influence of aromaticity appeared to be roughly inversely parabolic. Hence, increasing the number of aromatic rings increased bioavailability to a maximum after which there was negative effect on bioavailability. Although the effect of aromaticity appeared inversely parabolic, such a conclusion is only true for a compound containing up to four aromatic rings. In fact there seemed to be a point of inflexion when there were three aromatic rings present. If the number of aromatic rings was greater than four then
the proposed parabolic relationship may instead be similar to the relationship displayed for conformational minimum energy (Figure 6-28b). This nonlinearity in the relationship between descriptors and the target bioavailability necessitates any model to be trained on as broad a range of chemical structures as possible. The training set of compounds should adequately represent the chemical space of the test compounds to avoid the problem of extrapolation.

### 6.3.4 ANN Model Performance

The training, validation, and testing RMS errors for the optimum model with 10 descriptors were 19.21, 16.15, and 20.47 respectively. The strength of the correlation between selected descriptors and bioavailability corresponds to the quality of prediction. For training and validation subsets correlations between predicted and observed values of 0.736 and 0.897 respectively were achieved. These correlations further indicate that low predicted values correspond to the low observed bioavailability, and high predictions to high observed bioavailability values (Figure 6-29 and Figure 6-30).
Figure 6-29. Predicted bioavailability for training set of optimum ANN model.

Figure 6-30. Predicted bioavailability for validation set of optimum ANN model.
Chapter 6. QSPkR for a Diverse Series

For both training and validation subsets, more accurate predictions were made for compounds with observed bioavailability greater than 50%. Low bioavailability is most common with poorly water-soluble, slowly absorbed drugs. More factors affect bioavailability when absorption is slow or incomplete than when it is rapid and complete, hence, slow or incomplete absorption often leads to variable therapeutic responses. Many drugs have low oral bioavailability due to extensive first-pass metabolism.

Higher bioavailability values were predicted for the mainly basic drugs acyclovir, chlorothiazide, cromoglycate, domperidone, felodipine, fluorouracil, nimodipine, and sumatriptan. Bioavailability of acyclovir is dose dependent: absorption, and thus bioavailability, decreases with increasing dose, the nonlinearity of which may be difficult to account for in any model.

Negative bioavailabilities were predicted for alendronate, pamindronate, and selegiline during training. Since bioavailability can only take positive values then it is safe to assume that a negative prediction suggests extremely low clinical bioavailability. This observation is consistent with literature bioavailability of alendronate and pamindronate which both have reported values of 1% or less. Oral doses of selegiline are heavily metabolised on first-pass through the liver, and the active metabolites amphetamine and methamphetamine are produced. It is unclear whether clinical anti-depressant effects observed after oral administration are due to MAO-A inhibition by selegiline or the actions of amphetamine and methamphetamine. Clinical effect is dependent upon bioavailability so the low predicted value for selegiline may indicate a greater role of the active metabolites in producing the clinical effects seen with selegiline.
6.3.5 Independent Predictions

Predicted values for the independent validation set are shown in Figure 6-31 in ascending order of observed bioavailability. Where data ranges for experimental bioavailability were given in the literature, error bars were included to denote the range. Error bars were not included on predicted values because variations in intra-drug predictions were negligible. Again, more accurate predictions were made for compounds exhibiting higher experimental bioavailabilities such as cephalexin, lamotrigine, methyldopa, metoclopramide, nifedipine, paracetamol, propylthiouracil, and ranitidine. Predicted bioavailability for dolasetron correctly indicated that values were relatively high, and low bioavailability for losartan was correctly predicted.

Idarubicin, labetolol and naltrexone are all well absorbed rapidly after oral administration. Labetolol is extensively metabolised by the liver, and possibly in the gut wall, to O-phenyl-glucuronide, N-glucuronide and a glucuronide formed by conjugation at the secondary alcohol group. Once idarubicin is absorbed, it disappears rapidly from the blood and is distributed throughout the entire body. It shows a rapid distributive phase with a very high volume of distribution presumably reflecting extensive tissue binding. Naltrexone is subject to first-pass metabolism resulting in naltrexol and conjugated naltrexone and naltrexol as the major metabolites. Bioavailability of naltrexone varies greatly due to its substantial hepatic metabolism, with different studies reporting values between 5%–22% [Kogan et al., 1977; Meyer et al., 1984].
Figure 6-31. Optimum ANN model predicted vs observed bioavailability values for testing compounds (error bars denote experimental range).
Hydrochlorthiazide is absorbed throughout the small intestine and is not metabolised but is eliminated rapidly by the kidney with over 60% eliminated unchanged in the urine. Although relatively high, experimental bioavailability values span a large range which may explain the lower than expected predicted bioavailability for hydrochlorthiazide.

Overall, predictions were in good agreement with observed values for drugs exhibiting higher bioavailability. Predictions for these compounds generally lay close to or within the error range given for the observed values. Systemic bioavailability is a combination of absorption and metabolism of orally administered drugs. Compounds with high bioavailability would generally be well absorbed through the GI tract and not overly prone to first-pass metabolism either in the gut or by the liver. Compounds with lower bioavailability would either be poorly absorbed from the GI tract or substantially metabolised prior to becoming systemically available. Many physicochemical factors influence metabolic susceptibility of the compounds themselves, in addition to genetic and physiological characteristics of the human subjects. Enzymatic metabolism is a complex and diverse range of processes so compounds with poor bioavailability in the current study may not have been predicted well because of this complexity.

It is well known that all the information in an ANN model is contained in the weights connecting the neurons. Some researchers suggest that the ideal ratio of the number of training patterns to the number of connection weights, or $\rho$ parameter, lies within the range $1.8 < \rho < 2.2$. The claim is that ANN models with values of $\rho$ above this range may have insufficient connections to encode meaningful information, and models with values of $\rho$ below this range would have
too many connections and training data would then become memorised. Although
the optimum model in the current study had a $\rho$ value around 0.4, it has been
shown that ideal $\rho$ is implementation dependant and relies on the nature of the
training data itself (Chapter 3). Furthermore, the current study made use of both a
test set of compounds to examine model training, as well as an independent
validation set to examine predictive ability. Utilisation of data in such a manner
would virtually eliminate possible effects of memorisation. The optimum
predictive model was constructed with 31 neurons in the hidden layer. The
necessity of having such a large number of neurons and, hence, a large number of
connections indicated the inherent difficulty in modeling bioavailability.

The optimum model in the current study predicted higher than observed
bioavailability values for a number of compounds in the validation set, all of
which were reported to be well absorbed following oral dosing. Thus, for
screening purposes the current model may be suitable for compounds which have
been shown to be absorbed well from the GI tract. Alternatively, since for the
most part compounds were correctly predicted as having either higher (greater
than 50%) or lower (less than 50%) bioavailability, the model may be useful for
differentiating between compounds with either low or high bioavailability.

### 6.4 Results and Discussion – SWR Model

In Section 4.3.6.2 it was shown that multilinear regression models constructed
from descriptors selected by a nonlinear method resulted in worse models than the
original nonlinear model. In the present study instead of using descriptors chosen
by the nonlinear method, forward stepwise regression was utilised to select the best possible set of descriptors for a multilinear model.

### 6.4.1 Model Construction and Performance

The model constructed using stepwise regression contained eight descriptors (Equation 6-32):

\[
\text{Bioavailability (\%) } = -45.20 + 5.08 \text{ (El.aff)} + 4.09 \text{ (Ar.ring)} - 15.83 \text{ (HOMO)} - 3.34 \log P - 0.09 \\
\text{ (Molar.vol)} - 0.72 \text{ (Vol.HLB)} - 4.75 \times 10^{-7} \text{ (Water.sol)} + 1.18 \text{ (HH.bond.sol)}
\]

\textbf{Equation 6-32}

where HOMO is the highest occupied molecular orbital, log \( P \) is the calculated log \( P \) [Oxford Molecular, 2001], Molar.vol is the molar volume, Vol.HLB is the volumetric hydrophilic-lipophilic balance, Water.sol is the water solubility, and HH.bond.sol is the Hansen’s hydrogen bonding solubility parameter. The two remaining included parameters, El.aff and Ar.ring were the same as in the optimum ANN model (Section 6.3).

Training predictions for the stepwise regression model were close to observed experimental values for those compounds which had mid-range bioavailability (Figure 6-32). Predicted values were more deviated for compounds with relatively lower and higher observed bioavailability. Overall correlation of predicted values with observed bioavailability was 0.593 which explained 35% of the variance in the data. This represented the best model achievable using stepwise regression.
6.4.2 Independent Predictions

Predictions for the independent test set displayed a similar trend as for the training data (Figure 6-33). Predicted values were higher for compounds with low observed bioavailability and lower for compounds with higher bioavailability. In most cases the predicted bioavailability lay outside the experimental range associated with each compound. With the exception of hyoscine the stepwise regression model was able to qualitatively distinguish between compounds with bioavailability either greater than or less than 50%. Even so, the 10% error rate represented by hyoscine may be prohibitively expensive should this model be used for screening purposes.

---

Figure 6-32. Predicted bioavailability values for training set of SWR model.
Figure 6-33. Optimum SWR model predicted vs observed bioavailability values for testing compounds (error bars denote experimental range).

6.4.3 Descriptor Analysis

Two descriptors, El.aff and Ar.ring, were also included in the optimum ANN model. Thus, not all descriptors selected using nonlinear techniques are exclusive to nonlinear models. It is possible for descriptors to be useful in both linear and nonlinear models. The combination of descriptors in a nonlinear model is important though. Information can be gained not only from individual descriptors but from the information content of groups of descriptors in a nonlinear model.

Descriptors included in the stepwise regression model and their applicability to pharmacokinetics will now be discussed. The two descriptors included in both the ANN and stepwise regression models will not be further examined here.
Drug penetration through the biological membranes depends upon a number of molecular properties, such as lipophilicity, polarity, degree of ionisation and molecular size. On the other hand penetration between the cell gaps depends on molecular size and the concentration gradient. The octanol-water partition coefficient, $\log P$, is frequently used in QSARs as a measure of the lipophilic character of the molecules [Dearden, 1985]. Lipophilicity is approximately correlated to passive transport across cell membranes and the ability of a compound to partition through a membrane since membranes are composed largely of lipids [Brodie & Hogen, 1957]. Correlations of lipophilicity and membrane penetration have been extensively reviewed by Seydel and Schaper [Seydel & Schaper, 1981], and the role of lipophilicity has also been the subject of some recent work [Escribano et al., 1997]. Log $P$ is well established as a key parameter to describe lipophilicity, uptake and distribution in biological systems. However, $P$ is a ratio and a compound with low solubility in both octanol and water could have the same log $P$ as a compound with solubilities 100 times higher in both solvents [Egan et al., 2000]. Calculated log $P$ can be roughly correlated with drug absorption, but only for homologous series of compounds since it alone does not account for intramolecular interactions. For example, intramolecular hydrogen bonding can dramatically influence absorption properties [Norinder et al., 1999]. Therefore, other descriptors must be incorporated to account for these additional effects.

The steric effects characterise bulk properties of a molecule and can be described with molecular mass, surface area, density and molar volume. Small lipid insoluble substances penetrate cell membranes via the pores between aqueous
Chapter 6. QSPkR for a Diverse Series

phases on both sides of the membrane. The rate of such passive diffusion depends on the size of the pores, the molecular volume of the solute and the solute concentration gradient. Inclusion of Molar.vol in the final regression equation indicates the potential importance of diffusion processes in determining bioavailability of a drug.

Hydrophilic-lipophilic balance (HLB) is based on the concept that some molecules have hydrophilic groups, other molecules have lipophilic groups, and some have both. HLB is the ratio between the hydrophilic and the hydrophobic regions of a molecule, the balance of which describes which effect dominates in the molecule. Thus, percentage of each type of group in a molecule predicts what behaviour the structure will exhibit. The importance of lipophilicity \textit{in vivo} has received considerable attention. Parabolic or bilinear relationship between biological activity and chain length or partition coefficient are routinely observed [Hansch et al., 1968; Hansch & Dunn, 1972]. Poor intestinal permeability is associated with compounds that have low octanol/aqueous partitioning, contain strongly charged functional groups, high molecular weight, a substantial number of hydrogen-bonding functional groups, and high polar surface area.

Solubility parameters can be used to measure and predict the extent of interactions between materials. Aqueous solubility can be estimated using a group contribution approach [Klopman & Zhu, 2001]. This method allows an approximation of solubility to be made by calculating the contribution of relevant substructural units of the compounds. Since this method was developed with large boundary parameters and takes into account the interactions of many functional groups, it is well suited for calculating aqueous solubility for structurally diverse
Chapter 6. QSPkR for a Diverse Series

compounds. The total solubility of a drug can also be represented with the 3-parameter equation of Hansen [Crowley et al., 1966; Hansen, 1967], which was also included in the final regression model. The division of the solubility parameter into three components that take into account the dispersion, polar and hydrogen bonding effects of the solvent or solute considerably increases the accuracy with which non-ionic molecular interactions can be predicted, described and used to interpret solubility behaviour. One of the more useful aspects of this model is that the parameters can be visualised by plotting in a 3D space, with the hydrogen bonding, polar, and dispersion parameters forming the x, y, and z dimensions. For a particular drug, the resulting point in 3D space represents its solubility. Furthermore, a roughly spherical shape surrounding the point defines a “radius of interaction” for that drug.

Along with structural features and solubility characteristics encoded in various theoretical descriptors, electronic properties may also play a role in the magnitude of biological activity. HOMO energy is a useful descriptor that presents information on the distribution of $\pi$ electron and explains $\pi$-$\pi$ charge transfer interactions of unsaturated compounds. The high electronic density and high frontier orbitals are present in molecules with high electron delocalisation and can be used to predict biological reactivity. Increasing molecular reactivity also increases metabolic susceptibility. Therefore, higher reactivity is to be expected for the molecules with higher HOMO energies. HOMO energy plays a very important role in nucleophilic behaviour and it represents molecular reactivity as a nucleophile. Good nucleophiles are those in which electrons reside in high lying orbitals.
6.4.4 Comparison of ANN and SWR Models

Compared with the ANN model the stepwise regression equation did not perform well at prediction of bioavailability for independent compounds (Figure 6-31 and Figure 6-33). This was also a reflection of the poorer training results obtained using the stepwise regression equation. The inability of the stepwise regression model to represent the output bioavailability space was apparent given the low variance described by Equation 6-32 for the training set. Even though variance improved during the forward stepwise process, the maximum achieved using eight descriptors was still below 50%. Therefore, the linear combination of descriptors selected in Equation 6-32 was insufficient to make accurate predictions for independent compounds. On the other hand the ANN model trained substantially better than the stepwise regression model. Furthermore, a larger number of predictions by the ANN model lay within the experimental range of bioavailability than did predictions by the SWR model. The qualitative differentiation by the ANN model of test compounds with relatively high and relatively low bioavailability was also superior. More descriptors were required in the final ANN model which indicated that more subtle relationships between descriptors were developed by the ANN to encode the output bioavailability space.

An interesting point is that the ANN model trained on fewer compounds than the stepwise regression model but achieved better results. In general, the larger the training set the more reliable the model. The stepwise regression model was developed with greater than 10% more compounds compared with the ANN model. The reliability of the ANN model was superior to the stepwise regression
model since it was first validated after which testing was performed with a larger number of independent compounds than for the stepwise regression model. In this and most other aspects the ANN model proved superior to the stepwise regression model.

6.5 Conclusion

A QSPkR was successfully developed for a non-congeneric series of drugs using ANN modeling. The optimum model was trained on a structurally diverse range of compounds which enabled prediction of bioavailability for independent compounds not used in training. Such a model would be useful in screening potential drug candidates for suitable bioavailability characteristics.

Another model developed using regression techniques was examined and compared with the optimum ANN model. The regression equation was constructed using a larger number of compounds but was unable to describe the training data as well as the ANN model. ANN modeling proved more robust than multilinear regression.

Selection of optimal descriptors from amongst a large number allowed relationships between theoretical descriptors and bioavailability to be drawn. Thus, the relevance of certain theoretical descriptors to pharmacokinetics was further advanced.
Chapter 7

General Discussion and Conclusions
Chapter 7. General Discussion and Conclusions

This thesis has examined the use of artificial neural networks (ANNs) in developing quantitative structure-pharmacokinetic relationships (QSPKRs) in silico. Relationships were derived for a number of important human pharmacokinetic parameters. Drug data included a small congeneric series of β-adrenoceptor antagonists, a series of cephalosporins, and a large, structurally diverse set of compounds. The information required for the modeling process was obtained solely from drug structure without requiring in vitro or in vivo experiments. The conclusion from this thesis is that ANNs are a potentially valuable tool to aid in drug discovery and development.

This chapter details important findings of the work presented in this thesis. Key elements of the preceding chapters are discussed in a broader context and potential areas of further research are highlighted.

A summary of the major findings of this thesis is as follows:

1) The ratio of the number of training patterns to the number of number of connections, ρ, in an ANN model can affect model predictions. However, the ideal value of ρ is dependent upon the nature of the data itself and cannot be easily determined.

2) Complexity of ANN architecture reflects complexity of the pharmacokinetic parameter being modeled.
3) Simultaneous prediction of multiple pharmacokinetic parameters was able to be achieved by a single ANN model.

4) Useful information was gained from drug structure alone. Information varied from simple constitutional data to complex 3D descriptors.

5) Different theoretical descriptors conveyed different information for individual pharmacokinetic parameters.

6) Predictive QSPkR models were achieved for both congeneric and non-congeneric drug data sets.

7) ANNs were found to be a superior technique when compared with multilinear regression in QSPkR modeling.

A crucial aspect of data-driven studies including QSPkR analysis is the quality of the data itself. One method of obtaining the best possible data is to include an excess amount and then to select the most appropriate fraction. This is the rationale behind descriptor pruning. The selective method used in Chapter 3 and Chapter 4 was shown to be effective in constructing QSPkRs. It had the added advantages of permitting analysis and allowing complete control at every step. In the context of developing a tool for screening of large libraries of compounds it may be possible to construct a general QSPkR optimised to enable prediction of a single pharmacokinetic parameter. Development of such a model would require a large amount of computing and user-intercession time since permutations of a multitude of theoretical descriptors would need to be thoroughly investigated. Even though the potential benefits would be quite substantial, there is no
guarantee that a highly accurate general model would be attainable even after considerable expenditure of resources. Therefore, automated methods of descriptor selection are normally preferable.

Automated descriptor selection techniques are more suited to focussed or targeted libraries of compounds. They facilitate rapid construction of QSPkRs designed specifically for these less-general virtual chemical libraries. The benefit of speed using automated descriptor selection does outweigh the fact that they are less sensitive a method of descriptor selection than manual selective pruning. Automated methods of descriptor pruning have markedly improved over the past decade or so. Nevertheless, continued work in this area is required until a technique has been developed that is computationally inexpensive yet has the reliability of human control of the process.

In addition to constructing predictive QSPkRs, this thesis gave physicochemical and pharmacokinetic meaning to a number of theoretical descriptors. It is often questioned whether some theoretical descriptors have any rational meaning at all. The different influences that descriptors have on individual pharmacokinetic parameters demonstrated that a global meaning may not be possible for each descriptor. Rather, QSPkRs should be examined to determine the apparent meaning of a descriptor for each pharmacokinetic parameter that it has been related to.

In relating a descriptor to a pharmacokinetic parameter, the goal is to enable the synthesis of virtual compounds with structural characteristics that are optimised according to the important descriptors. The difficulty lies in the fact that many of
the 2D and 3D theoretical descriptors are complex and can take similar values for completely unrelated structures. Hence, moving from a descriptor value to a virtual structure is not a straight-forward process. This aspect of QSPkRs has not been explored to a great degree because of the complexity of such a task. Further research should be aimed at retro-assembly of virtual structures to suit a given QSPkR in order to design compounds with suitable human pharmacokinetics. Thus, instead of performing an passive screening process, the active assembly of optimised compounds may facilitate discovery of pharmacokinetically optimised potential drug candidates.

Using a model able to predict multiple pharmacokinetic parameters, optimisation of virtual structures would be even more comprehensive. This thesis has demonstrated the feasibility of developing such a model for a congeneric series of compounds. The number of models required to present a broader view of the pharmacokinetic behaviour of virtual compounds would be reduced by a model with multiple target outputs. Therefore, research aimed at developing a general model for structurally unrelated compounds would be of great benefit to the pharmaceutical industry.

It is very uncommon for there to be multiple target outputs in multilinear regression based QSPkRs. This is one additional advantage that ANNs have over regression models. The characteristic nonlinearity of ANNs further makes them more applicable to dealing with the complex space associated with chemical compounds. This fundamental difference between multilinear regression and ANNs is particularly apparent when QSPkR models are tested for predictive ability. It was shown in Chapter 6 that multilinear regression and ANN
approaches derived a different optimal set of descriptors from the same larger set, with the ANN model superior in both training and prediction. Moreover, regression equations developed in Chapter 4 from the optimum set of descriptors chosen by the ANN did not perform as well as the ANN model using the same descriptors. Therefore, nonlinear ANN modeling was the superior technique compared with multilinear regression for developing QSPkRs.

In conclusion, the utility of ANNs in pharmaceutics was demonstrated by successful construction of a number of QSPkRs. They have been shown to be a fast and reliable method for prediction of human pharmacokinetic parameters. ANNs have the potential to aid in drug discovery and development by providing a tool to complement existing screening techniques. It is not proposed that ANNs will replace current in vitro and in vivo screening tools. Rather, if they are appropriately incorporated into the overall drug design and development process they may provide considerable savings in resources and allow more rapid progression of potential drug candidates to the market.
References
References


Accelrys (2001). Cerius2 v.4.8 (software), San Diego.


ChemSW (2001). Molecular modeling pro demo v.4.07 (software), Fairfield.


CompuDrug International (1997). Pallas v.2.0 (software), Budapest.


References


References


References


References


References


MicroSimulations (1996). Accumodel v.1.0 (software), Mahwah.

Microsoft (1997). Excel v.8.0 (software), Mountain View.


References

MSI Molecular Simulations (1996). Catalyst v.3.1 (software), San Diego.


References


Schrödinger (2001). QikProp v.1.6.001 (software), New York.


References


References


Tripos (2000). Sybyl v.6.7 (software), St. Louis.


