Drugs of the Future 2004, 29(10): 1043-1057

CCC: 0377-8282/2004

**Review Article**

# Neural networks in ADME and toxicity prediction

***David A. Winkler***

*Centre for Complexity in Drug Design, CSIRO Molecular Science, Private Bag 10, Clayton South MDC 3169, Australia.*
*e-mail: dave.winkler@csiro.au*

**CONTENTS**



**Abstract**

The inability to understand and control the ADMET properties of molecules is an important reason why many candidate drugs fail late in the development pathway. Unfavorable pharmacokinetics, metabolism or toxicity, for example, can cause development candidates to be dropped. These failures are expensive and they contribute to the diminishing efficiency of the pharmaceutical industry. *In silico* models of ADMET properties allow these properties to be considered at an early, less costly stage, and should reduce the number of late-stage development candidates which fail. ADMET properties are multifactorial and complex, requiring very flexible methods to build predictive *in silico* models. This review summarizes the contribution neural networks are making to the development of useful ADMET models.

## Introduction

In a recent perspective article on the productivity of the pharmaceutical industry, Booth and Zemmel (1) examined six key reasons why the productivity of the industry has declined markedly over the last three decades. They identified one of the key causes of the decline to be poor chemical library design, resulting in poor pharmacokinetic profiles. It has become increasingly clear over the past decade that drug action is truly complex, and discovery paradigms that take into account the so-called ADMET (absorption, distribution, metabolism, excretion and toxicity) properties of candidate drugs are required. In earlier times, there was substantial focus on the target potency of drug leads and insufficient attention to how they will behave in a complex, biological *in vivo* system. Consequently, in the last 10-15 years intense effort has been focused on ways to determine experimentally how drug leads will behave in living systems. This has led to the development of rapid surrogate high-throughput assays for properties like phase I metabolism, blood-brain partition, protein binding, *etc.*

Major research efforts have focused on simulating ADMET properties of molecules. Clearly, if good *in silico* models were available for the important ADMET properties, it would be possible to focus early lead discovery into chemotypes with more “drug-like” properties, and reject those likely to be problematic in the later phases of drug development. As useful as rapid *in vitro* screening methods are, they cannot be applied when the leads are virtual, *i.e.*, designed or derived by computational means. Prediction of adverse ADMET properties can eliminate candidates before they are synthesized. The expectation is that *in silico* methods will reduce the large percentage of late-stage development compounds which fail due to ADMET properties, estimated to be as high as 90%. Given the amortized cost of getting a new drug to market of almost USD 900 million (2), any reduction in failure at the late, expensive end of the development process can generate large savings. Consequently, the modern drug discovery catchcry is “fail early, fail cheap”. Ekins summarized a suite of computational ADMET endpoints that are desirable in drug discovery (3). His primary models are solubility, absorption, mutagenicity, bioavailability, metabolic stability, blood-brain barrier (BBB) permeability, cardiac toxicity (hERG) and plasma protein binding.

## Modeling via a variety of approaches

ADMET processes are very complex and are difficult to model or simulate by computational methods. A wide variety of approaches must be adopted, ranging from

quite precise reductionist, 3-dimensional (3D) modeling studies of the interactions of drugs with various isoforms of cytochrome P-450 (CYP), through pharmacokinetic compartment modeling methods, to pattern recognition and artificial intelligence methods which view processes as complex systems and attempt to model their emergent properties. However, the focus of this review is the application of specific, model-free, nonlinear methods to modeling ADMET properties of small molecules such as hits, leads, drug candidates and drugs. We will summarize recent research that employs neural networks to this end. Such approaches are essentially pattern recognition and are useful for devising robust, predictive models of particular ADMET properties. In almost all cases, the methods attempt to find a relationship between the molecular properties and structures of drugs or drug leads and a specific ADMET property such as blood-brain partitioning. There are recent reviews on the application of neural networks to life sciences and to several other areas related to drug discovery and development, such as chemistry, structure-activity mapping, database mining, molecular diversity and combinatorial library design (4-11).

## Properties of neural networks

Neural networks are mathematical structures based broadly on the neural processes in the brain. Many types of neural networks have been devised and have been applied to a diverse selection of scientific and commercial applications. In the most general form they consist of a network of nodes and connections. The way the nodes are linked by the connections, the arrangement or architecture of the nodes and the nature of the transfer functions that the nodes apply distinguishes many of the types of neural networks. Two types of neural networks have dominated research into ADMET modeling thus far – the feed-forward, backpropagation neural network and the Kohonen net or self-organizing map (SOM). Backpropagation neural networks are essentially very general regression methods and are mainly used to develop quantitative models. Kohonen networks are a type of nonlinear mapping method and are used to cluster or classify molecules.

### *Backpropagation neural networks*

Backpropagation neural networks are “universal approximators” intrinsically capable of modeling any continuous function to an arbitrary degree of precision given suitable training data. They are model-free (not based on an internal model of a process), nonlinear, and are good at pattern recognition. These are excellent characteristics for modeling complex systems like living organisms, as they do not rely on understanding the processes “inside” in order to make good predictive models. They have been quite successful in building predictive models of *in vivo* biological responses.

Backpropagation neural networks (Fig. 1) are supervised regression methods that are trained using a data set in which the property to be modeled is known. The trained model can then be used to predict the required property for molecules not used in training. Information for each molecule in the training set is presented in turn to the input nodes and propagated via the connections to the hidden layer neurode processing elements. The weighted sum of the connections to each hidden layer node is processed by a nonlinear (usually sigmoidal) transfer function in each node, then passed on to the output node. Training occurs by comparing the response from the output node with the known response and propagating back the error through the network to provide a signal used to alter the weights and minimize the error.

Several authors have recently reviewed the application of backpropagation neural networks to drug discovery and development. Agatonovic-Kustrin and Beresford discussed potential applications of neural networks to the pharmaceutical sciences, ranging from interpretation of analytical data and drug and dosage form design, through biopharmacy to clinical pharmacy (12). Terfloth and Gasteiger reviewed how neural networks can be applied to rational drug design, gene prediction, locating protein-coding regions in DNA sequences, 3D structure alignment, pharmacophore perception, docking of ligands to receptors, automated generation of small organic compounds, and the design of combinatorial libraries (13). Ichikawa, as well as Yamashita and Takayama, have carried out mathematical analyses of neural networks and reviewed their application to optimization and prediction in pharmaceutical applications (14, 15). They proposed a

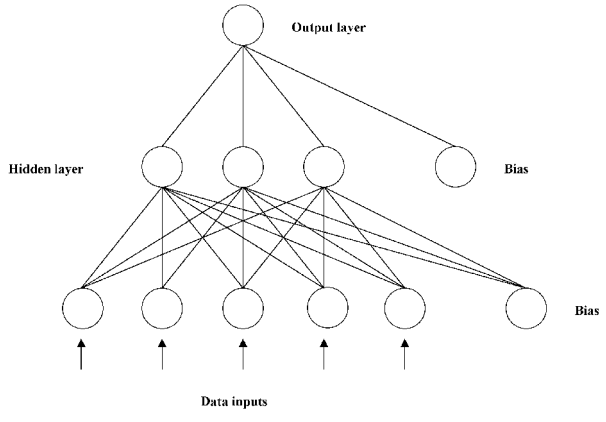


Fig. 1. A typical three-layer backpropagation neural network. Data for each molecule in the training set are applied in turn to the input nodes and propagated via the links to the hidden layer nodes. The weighted sum of connections to each hidden layer node is then modified by the transfer function inside the hidden layer node to produce the output of that node. The weighted sum of all hidden layer connections is applied to the output node. The transformed output of this node is compared with the known response variable (property being modeled) and the difference propagated back though the network to generate a correction to the connection weights, which reduces the error.

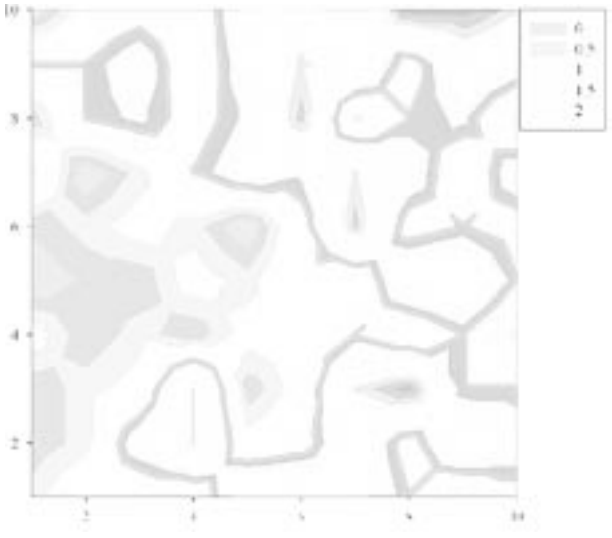

Fig. 2. A typical Kohonen net or self-organizing map (SOM). This example shows a Kohonen network trained with seven selected descriptors for cytochrome substrates. The map is color coded according to density of compounds and shows clustering of substrates and inhibitors into "islands". The SOM projects data from a higher dimensional space onto a 2D map, preserving the distances between objects. Although drawn flat here, the map is actually toroidal so that the top and bottom edges are joined, as are the left and right edges (from Ref. 91; reproduced with permission from the American Chemical Society).

descriptor-mapping method to find nonlinear relationships between the network outputs and descriptors. Winkler and Burden published a general review of the application of neural networks to combinatorial discovery and ADMET modeling (6).

*Self-organizing maps/Kohonen nets*

Self-organizing maps (SOMs), also called Kohonen nets, are another type of neural network whose main use is in clustering and classification. Kohonen mapping is an unsupervised procedure for comparing and classifying molecular data sets. Each chemical compound is represented as a point in the hyperspace defined by the molecular descriptors used. The SOM projects the multidimensional hyperspace onto a 2-dimensional (2D) map while preserving the order of distances between the points in a nonlinear way. An example of an SOM is given in Figure 2.

Gasteiger's group has done leading work on the application of SOMs to pharmaceutical research and has written several reviews summarizing the application of SOMs to pharmaceutical discovery and development (*e.g.*, 16). Polanski has employed a SOM network to map pharmacophores, discover similarity relationships and simplify the understanding of how descriptor hyperspace may influence drug properties (17). Kirew *et al.* have employed SOMs in automated data classification (18). These maps are often able to cluster compounds according to the mode of action or target of the drug. Zupan has compared the performance of a multibranching decision tree with a Kohonen neural network (19). The method is capable of mapping millions of multidimensional objects like spectra, structures, time series of process variables, multicomponent analyses of food or pharmaceutical products. Bayada and coworkers have used SOMs to distinguish between diversity and representativity (20), and Bernard *et al.* used SOM as an unsupervised procedure for comparing molecular databases (21). The aim of their work was to apply SOM to the study of the overlapping of two databases to obtain information about the extent of their molecular diversity. The results obtained indicated that SOM can be used for the search for new leads among available databases and the exploration of new structural domains for a given biological activity. Anzali and colleagues also reviewed the application of SOM to pharmaceutical discovery, with a focus on modeling of chemical libraries (22).

**Problems with neural nets**

Like all other regression techniques, neural networks can be overfitted (too many adjustable weights compared with the number of molecules in the training set) and can generate chance correlations (see below). They can also be trained for too long (overtraining), which results in them memorizing noise as well as signal in the data, with subsequent degradation in their predictive abilities. Most backpropagation neural networks avoid overtraining by stopping the training when the error in a separate "validation" data set reaches a minimum. Overfitting is usually avoided by using a neural network architecture (the number of hidden layers and numbers of nodes in each hidden layer) that is not too complicated. However, optimization of the neural network architecture within these constraints can still be a time-consuming, trial-and-error process. Manallack and Livingston reviewed the problems of neural networks in pharmaceutical research and concluded that they live up to their promise (23). They emphasized that one of the drawbacks of supervised learning methods is the danger of chance effects. They also investigated the potential for chance correlations in statistical treatments using neural networks. The group used random numbers as input data and found that networks were able to train successfully and reproduce the values of a random target. They proposed guidelines to minimize chance effects (24).

**Bayesian neural nets**

One way to improve the performance of backpropagation neural networks and avoid most of their problems is to use regularization. This involves adding a weight penalty term to the cost function that is minimized when the neural network trains. In unregularized regression the

cost function is simply the square of the error between the network or regression output and the known value for the output. By applying a Bayesian framework to this regularization process, one can find the optimum balance between variance (where an overly complex model fits the noise as well as the underlying relationship) and bias (where the model is too simple to capture the complexity of the underlying relationship in the data). Bayesian regularization produces the optimum, parsimonious model, in essence, automatically invoking Occam's Razor. An additional benefit is that training can be stopped when the evidence for the model is a maximum, obviating the need for a validation set. All available data can be used in the model.

Winkler and Burden first described the use of Bayesian regularized artificial neural networks (ANN) in quantitative structure-activity relationship (QSAR) modeling and drug discovery (25, 26). Shah *et al*. employed a Bayesian neural network to distinguish between drugs and nondrugs, using a commercial drug database as a surrogate for drug-like molecules and the Available Chemicals Database (ACD) as a surrogate for non-drug-like molecules (27). Their best results correctly predicted over 90% of the compounds in the drug database while classifying about 10% of the molecules in the ACD as drug-like. Bate *et al*. used Bayesian methods to analyze a database containing nearly 2 million reports of adverse drug reactions (ADR) held by the Uppsala Monitoring Center (28). They employed various computational approaches to extract signals from the database. A Bayesian confidence propagation neural network (BCPNN) was developed to manage large data sets and find drug-ADR combinations that are highly associated. The model was validated using a quarterly ADR update that identified 1,004 suspected drug-ADR combinations. Of these, 307 were potentially serious ADRs, 53 of which were related to new drugs.

## Training data and representation of molecular structure

### *Importance of data*

To build the most useful, predictive models of ADMET endpoints it is clearly important to have a large training set consisting of a diverse group of chemicals whose endpoints have been measured. Although the high-throughput screens for *in vitro* ADMET properties that have been developed by the pharmaceutical industry can generate large volumes of useful data, they are not accessible to those outside of the industry who are developing new modeling tools. Clearly there is some very good research into these areas in drug company laboratories, but the development of these methods and models is limited by the preferences and expertise of the individual researchers in the group. This nexus is one of the major impediments to the development of better ADMET modeling tools and predictive models.

### *Molecular representations (descriptors)*

Once a good training set is available, it is very important to know how to convert the molecules in the set into mathematical representations used in the modeling process. If this is done poorly, low-quality models will result. It is possible to calculate thousands of molecular descriptors, but many of these are not informative, or are highly correlated with other descriptors and contain similar information. Inappropriate choice of descriptors (mathematical representations of molecular properties), especially from a large pool of possible descriptors, can result in chance correlations arising, yielding models with little or no predictive ability. Fortunately, methods exist for determining the probability of chance correlations so they can be avoided with care (29).

Many descriptors were derived for QSAR modeling, a technique for building a wide variety of models of biological and physicochemical endpoints. Most of the work summarized here will involve the use of QSAR methods in one form or another. The QSAR method essentially involves finding a complex, usually nonlinear, relationship between the molecular properties of compounds in the training set and the biological or physicochemical property that each molecule exhibits in an assay or measuring system. As discussed below, neural networks are an extremely good way of building QSAR models. QSAR methods have been well described and reviewed in recent articles so a description will not be repeated here. Winkler and colleagues have written accessible reviews that introduce the concepts behind QSAR, point out problems that may be encountered, suggest ways of avoiding the pitfalls and introduce several exciting new QSAR methods discovered during the last decade (4-6).

Often simple molecular descriptors can contain quite rich embedded information that can be extracted by a universal approximator like a neural network. Burden reported how simple descriptors based on atom environment counts can give good models for physicochemical and biological target data (30). Winkler and Burden recently reviewed new developments in descriptor types (31). Until recently, QSAR analyses have used relatively simple descriptors based on substituent constants (*e.g.*, Hammett constants, pi or molar refractivities), physicochemical properties (*e.g.*, partition coefficients) and topological indices (*e.g.*, Randic and Weiner indices). New representations have been devised based on atom properties, eigen values of molecular matrices, E-state fields, topological autocorrelation vectors and various molecule fragment-based hash codes.

Gasteiger has reviewed the role of molecular descriptors in modeling of ADMET properties (32). Estrada has reviewed the use of topological indices in drug design and development, and summarized the most recent advances in this field (33, 34). Huuskonen *et al*. reviewed the role of the log of the partition coefficient between octanol and water (logP) as a useful parameter to correlate transport properties of drugs, model interactions between drugs and receptors, and map changes in the structure of drugs

with various biochemical or toxic effects (35). Hall and Kier introduced electrotopological state (E-state) indices for molecular structure description in which both electronic and topological characteristics are combined (36). E-state indices can be used to model many ADMET properties, including solubility and logP.

*Descriptor selection*

Several statistical methods have been used to select the appropriate molecular descriptors for model building. Traditionally, methods such as forward-selection/backward-elimination methods, principal components analysis and partial least squares (PLS) were employed. More recently, methods from artificial intelligence and complexity theory have been employed. Genetic algorithms have been a popular method of finding an optimum set of descriptors for modeling of properties. For example, Yasri and Hartsough developed a novel QSAR technique that used a genetic algorithm to preselect descriptors from a pool and neural networks that could dynamically modify its architecture to build optimum models (37).

However, many of these variable selection or reduction methods are linear. As the relationships being modeled are often nonlinear, a nonlinear variable selection or reduction method is preferable. Burden and colleagues employed a Bayesian method called automatic relevance determination (ARD) to prune out uninformative descriptors and to help interpret their neural net models (38). The ARD method ensures that irrelevant or highly correlated indices used in the modeling are ignored and gives a relative weighting for the importance of each descriptor to the model. Agrafiotis and Lobanov reviewed dimensionality reduction techniques from the statistical literature, multidimensional scaling and nonlinear mapping to reproduce the topology and structure of the data space in a faithful and unbiased manner (39). Because these methods can be computationally demanding, these authors developed a novel approach that combines conventional nonlinear mapping techniques rooted on the principle of probability sampling, with feed-forward neural networks to build models orders of magnitude larger than those accessible with conventional methodologies.

*Predictive versus interpretive models*

The focus of model building can be interpretation or prediction. Interpretation involves dissecting the model to learn more about the interaction of a set of molecules with a target. Interpretation is facilitated by data of very high quality, descriptors that are easily mapped back into molecular features familiar to synthetic chemists, and a mapping method in which the form of the model and the contribution of each descriptor to the model are transparent. Predictive models often involve much larger training sets, often of lower quality (from high-throughput screens, for example), computationally efficient descriptors and less transparent mapping methods. Predictive methods aim to give the best estimate of a property for a molecule not yet synthesized and to assist in prioritizing lead discovery. Most of the models discussed here will be predictive models that have been generated using neural networks.

## Application of neural networks to modeling specific ADMET properties

*Drug-likeness*

One of the first and most widely accepted rules for accounting for ADMET properties was the “Rule of Five” proposed by Lipinski *et al*. (40). The success of this paradigm stimulated research into factors distinguishing drugs from nondrugs. Neural networks have been employed in several recent studies of drug-likeness. Walters and Murcko reviewed recent developments in combinatorial chemistry and high-throughput screening and automated methods of determining which compounds from a library should be synthesized and screened (41). Methods range from simple counting schemes to sophisticated machine learning techniques such as neural networks. While many of these methods perform well in validation studies, the field is still new.

Murcia-Soler *et al*. used topological and structural descriptors and a backpropagation neural network to discriminate general pharmacological activity, including drug-like propensity (42). Sadowski and Kubinyi developed a scoring scheme for the classification of molecules into drugs and nondrugs (43-45). They used large databases of drugs and nondrugs, atom type descriptors, and trained a backpropagation neural network to classify molecules. Their methods correctly classified about 80% of compounds in sample databases and allowed the drug character of combinatorial libraries to be optimized. Frimurer *et al*. also used a backpropagation neural network, 2D descriptors and the MACCS Drug Data Report (MDDR) and ACD databases to classify compounds into “drug-like” and “nondrug-like” classes (46). The method correctly assigned 88% of the compounds in both MDDR and ACD and gave a much better prediction performance than the “Rule of Five”, which accepted 74% of the ACD compounds and only 66% of those in MDDR. Bruestle and coworkers derived a set of descriptors from semiempirical MO (AM1) calculations and used the World Drug Index and Maybridge databases as surrogates for drug-like and nondrug-like compounds (47). They trained a Kohonen net for the entire Maybridge data set, and then projected the drug database onto the resultant map, resulting in a clear distinction between drugs and nondrugs and also between hormones and other drugs. In another study, Shah *et al*. used two different sets of 1-dimensional (1D) and 2D descriptors and a Bayesian neural network to distinguish between drugs and nondrugs (27). Their models correctly predicted over 90% of the compounds in the Comprehensive Medicinal Chemistry database to be drug-like, while classifying

about 10% of the molecules in the ACD as drug-like. They tested the generalization ability of the models using the MDDR database and predicted roughly 80% of the molecules in the MDDR to be drug-like. They proposed using the models to design combinatorial libraries.

Zernov *et al*. used another machine learning method called Support Vector Machines (SVM), a powerful classification and regression tool, to determine drug-likeness (48). They compared SVM with well-known neural network techniques to predict drug-likeness and agrochemical-likeness for large compound collections. For their data set they found that SVM outperforms various neural networks using the same set of descriptors. Similar results were reported by Byvatov *et al*., who also used SVM and ANN to classify compounds into drug/nondrug classes (49). They found that the SVM classifier yielded slightly higher prediction accuracy than ANN, independent of the type of descriptors used, the size of the training data sets, and the algorithm employed for neural network training. Although SVM outperformed the ANN classifiers with regard to overall prediction accuracy, both methods were shown to complement each other and the authors suggested that a consensus approach would yield better classification than either method individually. They also reviewed the theory of SVM and neural network training. Takaoka *et al*. adopted a novel method of distinguishing drug-like compounds by capturing the collective experience of working chemists using a neural net and SVM (50). Five chemists assigned a drug-like and "ease-of-synthesis" score to each of 3,980 diverse compounds. The resulting models were found to efficiently eliminate compounds that were not drug-like and/or hard to synthesize.

## Solubility, partition and permeability models

The water solubility, logP and ability to penetrate biological barriers such as the blood-brain barrier (BBB) or intestinal mucosa are important properties for drug candidates. Neural networks have been very effective in building robust, predictive models for these properties, although published work has suffered from a paucity of training data for the latter two properties. Several general reviews of solubility and partitioning modeling have recently appeared. Taskinen and Yliruusi reviewed the literature describing neural network modeling of physicochemical properties of compounds from molecular structure. Properties such as logP, water solubility, boiling point and vapor pressure have been modeled by several research groups using different approaches and structurally diverse large training sets. In most cases, the prediction accuracy of models was close to the measurement accuracy (51).

### *Solubility*

Eros and colleagues presented a comprehensive review of water solubility prediction methods. Multiple linear regression (MLR) analysis, PLS and neural networks have all been employed to model aqueous solubility (52). Standard errors of prediction from neural network models were typically 0.72 log S units. Huuskonen *et al.* used neural networks to model the aqueous solubility of 211 drugs and related compounds representing acidic, neutral and basic drugs of different structural classes (53). Structural parameters used as inputs to the neural network included electrotopological indices and topological indices. They achieved a predictive $r^2$ value (squared correlation coefficient between predicted and measured outputs) of 0.86 and a standard error of prediction of 0.53 log S units for this smaller data set. Ran and coworkers compared the revised general solubility equation (GSE) with neural networks and MLR methods for their ability to model aqueous solubility (54). They found the GSE and ANN predictions to be more accurate than MLR methods. Although the GSE used only two parameters and no training set, its average absolute error was only 0.1 log units larger than that of the neural net model. Liu and So reported a simple aqueous solubility model based on seven 1D and 2D descriptors and a neural network (55). The model achieved a prediction error of 0.72 log S units for a diverse set of 1,312 organic compounds, comparable with the estimated experimental uncertainty of no less than 0.5 log S units. Manallack and coworkers employed BCUT (Burden, CAS and University of Texas) descriptors (eigenvalues of modified molecular connectivity matrices) to discriminate compounds with poor aqueous solubility (56). Approximately 95% of compounds were classified correctly using this filter. Bruneau employed a very diverse data set consisting of literature and proprietary compounds and a Bayesian neural network, to build a robust model of solubility (57). About 100 descriptors emphasizing surface properties were used in the models. The importance of the descriptors to the models was established by means of a modified Gram-Schmidt or ARD procedure. Todeschini's research group has developed a comprehensive computer program for personal computers called DRAGON which is capable of calculating almost 2,000 common molecular descriptors useful for prediction of solubility and other properties (58).

### *LogP prediction*

LogP was one of the earliest physicochemical properties of molecules found to be important in drug action and ADMET properties. It is one of the most extensively modeled properties, given its importance in many biological systems.

Eros *et al.* critically reviewed the published methods of logP prediction and compared the predictive power of commercial software packages and their recently developed automatic QSPAR program (59). They trained their models on a very diverse set of 625 known drugs and drug-like molecules with experimentally determined logP values and 78 "outliers", compounds that were not well predicted by traditional methods. Tetko reported modeling logP using an associative neural network, a combination of an ensemble of backpropagation neural networks and

k-nearest neighbor clustering techniques (60). An associative neural network has a memory, and if new data become available, the network improves its predictive ability without the need to retrain the neural network ensemble. According to Tetko, this feature of the method dramatically improves its predictive ability over traditional backpropagation neural networks. Devillers *et al*. also employed a backpropagation neural network and a training set of 7,200 chemicals to build a predictive model of logP (61). Their most predictive model had excellent statistics, with a training set root mean-squared error (RMSE) of 0.37 log units and an $r^2$ of 0.97, and a test set RMSE of 0.39 and $r^2$ of 0.98.

Huuskonen *et al*. reported several studies evaluating the efficacy of atom-type E-state indices for logP prediction using large, diverse training sets (62). They compared MLR analysis and neural networks as mapping tools to generate logP models and found that both methods provided reliable logP estimates, although neural networks provided better prediction ability for training and test sets. For the best models, $r^2$ and cross-validated RMSE were 0.90 and 0.46 log units, respectively. When the model was used to predict logP values for a completely independent test set, a predictive $r^2$ of 0.94 and RMSE of 0.41 log units were obtained. The authors attributed the improved predictive ability of ANN to the nonlinear properties of this method. Schaper and Rosado Samitier also used neural networks to recognize structural features important for logP models (63). Using a three-layer network with three neurons in the hidden layer and indicator (binary) variables to indicate the presence or absence of atom types and bond types in molecules, a standard error of 0.25 log units was obtained for a small training set. For a test set of 50 similar compounds, the standard error of prediction was 0.66 log units.

*Blood-brain barrier (BBB) partition*

Predicting BBB partitioning is clearly important for targeting new drug candidates to the central nervous system (CNS) for neurological targets, or away from the CNS if the candidates have peripheral targets and CNS effects are to be avoided. The ability to build robust, predictive models of BBB partitioning has been hampered by the paucity of experimental data to train models. Consequently, all reported models are based on small, usually very similar, training sets with limited predictive power.

Doniger and colleagues used two different machine learning algorithms, a backpropagation neural network and a support vector machine (SVM), to predict the BBB permeability of different classes of molecules (64). Both algorithms were trained on a data set of 179 CNS-active molecules and 145 CNS-inactive molecules. Molecular descriptors included molecular weight, lipophilicity, hydrogen bonding capacity and other variables likely to modulate the ability of a molecule to diffuse through a membrane. In this classification problem (permeable/not permeable), the SVM outperforms the neural network, correctly predicting an average of 81.5% of molecules in the test sets, compared with 75.7% for the neural network. Fu *et al*. also used a neural network model to quantitatively predict log BBB from structural parameters such as molecular volume, atomic charges of oxygen and nitrogen hydrogen bond acceptors (65). For a training set of 56 compounds and a test set of 5 compounds, RMSE values of 0.24 and 0.26 log units were obtained, respectively. Winkler and Burden used a Bayesian neural network to build a robust model of BBB partitioning using a data set of 106 diverse compounds (66). They compared three types of molecular descriptor – property-based, topographical and eigen value – for efficacy in modeling log BBB. The results showed that the property-based descriptors were best for training the model, consistent with other reported observations. However, all three classes of molecular descriptor had similar abilities to predict log BBB of an independent test set. The models could account for approximately 65% of the variance, with the remainder likely to be experimental error. Standard errors of prediction were 0.55 log units.

*Human intestinal absorption (HIA)*

The ability of a molecule to be absorbed through the human intestinal cell lining is an important property for potential drug candidates. Measuring this property is costly and computational models that estimate percent human intestinal absorption (%HIA) are attractive alternatives. The ability to develop better modeling tools for HIA and to build robust models is also limited by the lack of large, diverse data sets in the public domain. Podlogar and Muegge have published a review of computational methods for the estimation of intestinal absorption (67).

Agatonivic-Kustrin and coworkers developed a neural network-based model of %HIA using 86 drug compounds and their experimental intestinal absorption values (68). They used constitutional, topological, physicochemical, geometrical and quantum chemical descriptors. A supervised neural network with radial basis transfer function was used to build the model. A genetic algorithm was used to select the most important descriptors so that overfitting was controlled. Their 15-descriptor neural net model had a training set RMSE of 0.59 and a test set RMSE of 0.90. The results suggested that lipophilicity, conformational stability, polarity and hydrogen bonding have the largest impact on intestinal absorption. Wessel *et al*. also studied a set of 86 drug and drug-like compounds with measured values of %HIA taken from the literature and developed a model (69). They used a neural network coupled with a genetic algorithm to determine the best descriptors and network architecture. The best model had an RMSE of 9.4 %HIA units for the training set, 19.7 %HIA units for the cross-validation set and 16.0% HIA units for an external test set. Niwa used 2D molecular descriptors and two variants of a radial basis function neural network to model %HIA in the same data set of 86 compounds studied by Wessel (70). The radial basis

function variants used were the regression neural network and a probabilistic neural network, both of which performed well in modeling %HIA values. The RMSE was 22.8 %HIA units for an external test set for the regression neural net model and 80% of the external test set was correctly classified for a probabilistic neural network model.

Fujiwara *et al.* used a combination of descriptors derived from quantum chemical calculations and a neural network to predict Caco-2 cell permeability from structure (71). For a training set of 87 compounds, the dipole moment, polarizability, sum of charges on nitrogen and oxygen atoms, and hydrogen atoms bonding to heteroatoms were calculated as descriptors. A backpropagation neural network was used to build the model for predicting Caco-2 cell permeability. A cross-validation procedure revealed that the neural network model had good predictability, with a predictive RMSE of 0.507 log units compared with RMSEs of 0.584 and 0.568 log units for linear and quadratic regression models, respectively. In a similar study, Gohlke and colleagues employed a combined simulated annealing/neural network approach to derive a model for human intestinal absorption for drugs and drug-like compounds with an overall error of prediction in the range of the experimental error (72).

**Pharmacokinetic/pharmacodynamic models**

As Saxena and Schaper have pointed out, the quantitative analysis and physicochemical description of time- and dose-dependent *in vivo* drug effects are problematic (73). The observed effects depend on both pharmacodynamics and pharmacokinetics, factors that are influenced in different ways by physicochemical properties of drugs. *In vivo* effects of drug series are governed by highly nonlinear, extremely complicated and often unknown relationships. Neural networks are good candidates for modeling these types of properties. Gobburu and Chen have summarized the advantages of the flexible, nonlinear, model-independent properties of neural networks in modeling pharmacokinetic (PK)/pharmacodynamic (PD) data (74). Neural networks are flexible enough to accurately predict PD profiles for a wide variety of PK/PD relationships and can accurately predict PD profiles without requiring any information regarding the active metabolite.

Saxena and Schaper studied *in vivo* hypotensive effects of drugs in cats using a neural network to model the unknown nonlinear relationship between input variables and observed effect values (73). Although each drug was applied at only two doses and the effect was observed at only a few time points, the neural network was able to predict the complete time-activity profiles (PK) for the drugs. However, complete dose-response curves could not be obtained due to insufficient applied dose data. Opara *et al.* also used neural networks to predict PK responses of individuals in drug bioequivalence studies (75). Efficient prediction of the responses allows identification of deviations from typical population values. The neural net model prediction errors of PK parameters for verapamil were comparable to the observed differences between formulations in the bioavailability study. In another neural network modeling study, this same group used data from three distinct bioequivalence studies of oral verapamil products involving a total of 98 subjects and 312 drug applications. The average absolute prediction errors for the PK parameters area under the concentration-time curve (AUC), peak plasma concentration ($C_{max}$) and time to reach peak plasma concentration ($t_{max}$) were 30.5%, 39.6% and 30.7%, respectively (76).

Hussain explored the properties of neural networks for predicting *in vivo* drug concentration-time profiles and for multidimensional interspecies scaling of PK parameters. These models provided adequate generalization although they were trained with limited examples. The availability of animal data in at least four different species was a major limitation in the modeling (77).

Turner and coworkers used radial basis functions and theoretical descriptors to develop a quantitative structure-PK relationship for structurally diverse drugs (78, 79). All models were trained on data from 137 compounds, tested with a set of 15 compounds and evaluated for predictive ability with an additional 15 compounds. The PK parameters modeled were clearance, fraction bound to plasma proteins and volume of distribution. The best model had training and test set $r^2$ values of 0.736 and 0.897, respectively. In general, predictions from the model agreed well with experimental values.

Veng-Pedersen and Modi reviewed the role of neural networks in modeling biological systems (80). They pointed out their particular suitability for dealing with PK and PD systems, especially in cases of multivariate PK/PD population kinetics when the systems are so complex that modeling by a conventional structured model-building technique is very difficult. They demonstrated the application of neural network modeling to PD by predicting the CNS activity of alfentanil. The relative prediction performance of the neural network was 66%, which indicates an excellent prediction given the intrinsic fluctuations in the effect variable. Haidar *et al.* used neural networks to develop a predictive population PK/PD model for repaglinide, an oral hypoglycemic agent (81). PK/PD and demographic data from a dose-ranging phase II trial were divided into a training set and test set. PK and PK/PD models were constructed and compared to naïve averaging and randomly generated numbers. These authors concluded that neural networks offer a quick and simple method for predicting PK and PD properties, for identifying significant covariates and for generating hypotheses.

Ritschel and Akileswaran used a backpropagation neural net with a combination of physicochemical properties to model animal PK parameters (82). Fourteen network models, using a variety of input variables, were developed. Protein binding, partition coefficients, dissociation constants and the total clearance and volume of distribution of 41 drugs were measured in rats and dogs and then used for prediction of human total clearance and volume of distribution. Drugs with a logP < 1.17 showed

a predictability of 63.4% for total clearance and 41.5% for volume of distribution.

Moon and Smith developed a preliminary PD model for dosing of the hydroxymethylglutaryl CoA-reductase inhibitors simvastatin and atorvastatin using a neural network (83). Lipid panels from 17 patients were used as inputs to the model, and doses of simvastatin or atorvastatin that achieved those lipid results were the outputs. The dose predictions were compared with the actual doses given by the hospital. The neural network model based for one data set predicted a dose that was within 95% of the actual dose 7 of 12 times and predicted use of the drug actually used 13 of 19 (68.4%) times. The model based on a second data set was less successful, predicting that the drug was actually used 10 of 17 (58.8%) times, but the predicted doses were always substantially less than the dose actually used. The neural network model for the dosing of simvastatin and atorvastatin could predict appropriate dosing but inclusion of other factors (*e.g.*, age, body weight, sex) and a larger sample size may be necessary for development of a more accurate model.

Wajima *et al*. compared the performance of three types of regression methods – MLR, PLS and neural networks – to predict human clearance of drugs (84). The training set consisted of clearance data for rats, dogs and humans for 68 drugs. Molecular weight, logP and the number of hydrogen bond acceptors were used as descriptors. Schneider *et al.* also used several statistical regression models and neural networks to predict the hepatic drug clearance in humans from *in vitro* (hepatocyte) and *in vivo* PK data (85). The results indicated that human hepatocyte data was the best predictor, followed by rat hepatocyte data. They concluded that the most cost-effective and accurate approach to achieve satisfactory predictions in humans is to use a combination of *in vitro* clearance in human and rat hepatocytes. The main limitations of the approach were related to the limited amount of experimental data available.

## Metabolism modeling

Human hepatocyte high-throughput assays are now available to screen drug candidates efficiently, providing a good source of *in vitro* data for modeling of phase I metabolism. Similar high-throughput methods are not as well developed for phase II metabolism. However, modeling the metabolic fate of compounds is very difficult, partly because of the combinatorial explosion of possible metabolites that occurs in typical decision tree- and rule-based systems.

Holmes *et al*. reported that high-resolution proton NMR spectroscopy of biofluids and tissues coupled with backpropagation and probabilistic neural networks can provide complementary data for use in *in vivo* toxicological screening of drugs (86). The authors used NMR spectroscopy to characterize the time-related changes in the urinary metabolite profiles of rats treated with 13 model toxins and drugs which predominantly target the liver or kidney. The data consisted of 1,310 samples, of which 583 were used in a training set and the remaining 727 in a test set. The probabilistic neural network model gave superior results to the backpropagation neural network model, distinguishing 13 classes of toxicity with better than 90% accuracy.

### *Phase I metabolism*

Vermeulen reviewed the role of CYP in important biotransformations of drugs and other xenobiotics (87). The paper summarized various computational approaches used to rationalize and predict the activity and substrate selectivity of CYP, as well as the possibilities and limitations of these approaches. Since human CYP2D6 is one of the most important drug-metabolizing enzymes, this isozyme was chosen as a focus of the review. The authors concluded that no one computational approach is capable of rationalizing and reliably predicting metabolite formation by CYP2D6, and a consensus of several methods is therefore desirable.

Zuegge and coworkers reported the development of computational PLS- and neural network-based prediction systems for binary classification of drug-drug interaction liability caused by CYP3A4 inhibition (88). The system was trained using $IC_{50}$ values for 311 molecules. The best model correctly predicted 95% of the training data and 90% of a semi-independent validation data set. This group also reported a study of the comparative prediction accuracy of several different mathematical models for human hepatic metabolic clearance: allometric scaling, physiology-based direct scaling, empirical *in vitro-in vivo* correlation and supervised neural networks (89). The study used a publicly available data set of 22 extensively metabolized compounds. The latter three modeling approaches yielded $r^2$ values > 0.77, compared with $r^2$ values < 0.44 using the other method. The percentage of successful predictions ranged from 55% to 68%.

Molnar and Keseru also adopted a neural network approach to identify potential CYP3A4 inhibitors (90). They used training data from the Genetest database and 2D molecular fingerprints as descriptors. The neural net model correctly identified at least 89% of CYP3A4 inhibitors in the validation set. Korolev *et al.* used a Kohonen SOM to study CYP-mediated metabolic transformations of xenobiotic molecules (91). They compiled a database of 2,200 compounds comprising known human CYP substrates, products and nonsubstrates for 38 enzyme-specific groups. They determined the CYP-mediated metabolic reactions most typical for each group and examined the substrates and products of these reactions. They used an SOM neural network together with physicochemical descriptors to produce maps visualizing isozyme-specific groups of substrate molecules.

Vermeulan has suggested that when effective links with other new and recent developments such as bioinformatics, neural network computing, genomics and

proteomics are more mature, *in silico* rationalization and prediction of drug metabolism by CYPs is likely to become one of the key technologies in early drug discovery and development processes (87).

### *Phase II metabolism*

Modeling of phase II metabolism is much less established than for phase I, due partly to the lack of adequate data sets. Sorich *et al.* conducted a modeling study of 12 isoforms of human UDP-glucuronosyltransferase (UGT), an enzyme superfamily involved in the metabolism of drugs, nondrug xenobiotics and endogenous compounds (92). They compared PLS discriminant analysis, Bayesian-regularized neural networks and SVM methodologies for their ability to classify substrates and nonsubstrates. Simple 2D descriptors were used to characterize molecules in the training and test sets. The best models showed predictive ability for all UGT isoforms, with 5 of 12 isoform test sets exhibiting > 80% prediction accuracy. In a recent paper, this group reported an extension of this work in which quantum chemical density functional theory descriptors related to the mechanism of the UGT enzyme superfamily were approximated by electronegativity equalization methods (93). SVM methods combined with a consensus approach yielded substantial improvements in the predictive ability of the glucuronidation models. The success of pattern recognition methods in modeling this major phase II metabolic process should lead to further research activity by computational scientists in this important area.

## Modeling of toxicity, mutagenicity and carcinogenicity

There has been significant activity in the application of neural networks and other artificial intelligence methods to toxicity modeling, driven in large part by environmental toxicology. Schultz and colleagues reviewed QSAR approaches to modeling toxicology in both environmental and human health areas since 1995 (94). They compared the rules-based expert systems with self-organizing dynamic algorithms such as neural networks. They also summarized the current status of modeling human health effects, including mutagenesis and carcinogenesis, developmental toxicity, skin sensitization, and skin and eye irritation.

### *Cardiac toxicity (hERG)*

Agents that cause hERG (human ether-a-go-go related gene) channel blockade are a major concern in drug design, as they can cause sudden cardiac death. Roche *et al.* have reported a computer-based method for prediction of the hERG potassium channel affinity of organic compounds (95). They applied several techniques to modeling SAR in hERG: substructure analysis, self-organizing maps, principal component analysis, PLS and supervised neural networks. The neural network model was the most accurate prediction system, correctly classifying 93% of the nonblocking agents and 71% of the hERG channel blockers. The authors advocated using the neural network model as a virtual screening method for chemical library focusing and combinatorial library design.

### *Acute toxicity*

The measurement of toxic effects is time-consuming and expensive, providing an impetus to develop fast computational methods to predict toxic effects from the molecular structure. Predicting toxic effects is challenging because there are usually multiple toxic mechanisms involved. Feng and colleagues compared several combinations of different chemical descriptors and popular statistical methods for their efficacy in predicting toxicity (96). They employed recursive partitioning, neural networks and partial least squares to model the response surface. All of the methods and descriptors worked to a degree, but the authors suggest that certain descriptors work better with specific statistical methods than with others, indicating the need for the development of better methods.

### *Mutagenicity/genotoxicity*

Valkova *et al.* reported the application of a counterpropagation neural network to modeling the mutagenicity of 95 aromatic and heteroaromatic amines collected from the literature (97). The molecules were represented by topological, geometrical and quantum chemical descriptors. Predictive ability of the final model was excellent, with training and test set $r^2$ values of 0.98 and 0.82, respectively. Validity of the best model was confirmed by a randomization test. The authors suggested that neural networks are powerful tools for modeling structure-mutagenicity relationships. Vracko *et al.* also analyzed a data set of 95 aromatic amines for their mutagenic potency using a counterpropagation neural network, and closely related descriptors, to develop models to predict mutagenicity (98). In this case, the models produced $r^2$ values between 0.65 and 0.75, comparable with models obtained by linear methods. The authors also reported an analysis of the data using SOM, which identified clusters of structurally similar compounds. In a related study, this group modeled 12 trimethylimidazopyridine isomers with varying mutagenic potency toward 2 strains of *Salmonella* (99). Quantum chemical and 3D descriptors and a counterpropagation neural network were used to build the models. They reported predicted mutagenicities for two isomers that have not yet been synthesized.

Karelson and coworkers modeled Ames test genotoxicity of aromatic and heteroaromatic amines using the Chebyshev polynomial expansion and neural networks

(100). They used quantum chemical molecular descriptors to compare the nonlinear models with linear models from MLR analysis and to determine the chemical features that most influence mutagenicity.

Bahler *et al.* reported a preliminary study that used rodent carcinogenicity data to build *in silico* models aimed at supplementing the costly and time-consuming laboratory tests for carcinogenicity (101). Their models took the form of decision trees, rule sets, neural networks, rules extracted from trained neural networks and Bayesian classifiers. Their training set was obtained from rodent carcinogenicity bioassays conducted by the National Toxicology Program on 226 test articles. Descriptors used included physicochemical parameters, structural alerts, *Salmonella* mutagenicity assay results, subchronic histopathology data, and information on route, strain and sex/species for 744 individual experiments. They concluded that no single carcinogenicity model would likely be adequate and that a consensus of several methods would give the best predictions. Brinn *et al.* built a neural network-based classification model from a large, structurally heterogeneous data set of mutagens and nonmutagens (102). Substructural data composed of 1,280 fragments were used as inputs. The best models misclassified only 11% of the test set and 6% of the training set. The authors analyzed the output of the neural network model's hidden layer to discern clusters of mutagens and nonmutagens, demonstrating how the network was classifying the data.

Lindquist and colleagues reported a method of detecting new drug safety signals using an international database of more than 2 million case reports from the World Health Organization Program for International Drug Monitoring (103). They developed a new Bayesian logic data mining signaling process, implemented using a BCPNN, to aid clinical review of the data. The study tested the predictive value of the BCPNN in new signal detection by comparison with literature sources and with an established signaling procedure. The BCPNN method detected signals with a positive predictive value (probability that a positive prediction is correct) of 44% and a negative predictive value (probability that a negative prediction is correct) of 85%. The authors concluded that the BCPNN approach had a high and promising predictive value in identifying early signals of new ADRs.

## Delivery

There has been recent and promising activity in the application of neural networks to formulation and controlled-release problems. Although outside the scope of this review, a brief summary of key review articles is provided.

### *Formulation*

Agatonovic-Kustrin *et al.* developed a colloidal dosage form for the oral delivery of rifampicin and isoniazid in combination with the aid of neural network data modeling (104). Data from 20 pseudoternary phase triangles containing miglyol 812 as the oil component and a mixture of surfactants or a surfactant/cosurfactant blend were used to train, test and validate a neural net having radial basis function network architecture. The best model successfully predicted the microemulsion region as well as the coarse emulsion region, but failed to predict the multiphase liquid crystal phase for cosurfactant-free systems. A novel microemulsion formulation capable of delivering rifampicin and isoniazid in combination was created, and the model assisted in understanding the process of microemulsion formation and stability within pseudoternary colloidal systems. McCall and coworkers have also employed neural networks as a tool in the iterative process of formulation development (105). They used the neural network to develop relationships between formulation variables, observed *in vitro* dissolution data, and simulated plasma drug concentration-time profiles.

### *Controlled release*

Controlled-release drug delivery systems offer important advantages over conventional dosage forms. However, there are great challenges in efficiently developing controlled-release drug delivery systems due to their complexity. Traditionally, a statistical response surface method is employed to develop and formulate controlled-release dosage forms. Sun *et al.* have reviewed the application of neural network methods to the design of controlled-release drug delivery systems (106).

Reis and colleagues used neural networks to design two controlled-release systems, namely hydrocortisone in a biodegradable matrix and rhodium(II) butyrate complexes in a bioceramic matrix (107). The models simulated release profiles as a function of fundamental properties such as diffusion coefficient, saturation solubility, drug loading and the height of the device. The neural network could essentially quantitatively predict ideal experimental outcomes and this approach was found to be useful for the efficient design of controlled-release systems. Lim and colleagues also used a neural network-based intelligent learning system for the prediction of drug release profiles in transdermal iontophoresis (108). They employed a Gaussian mixture model to model experimental data and predict the drug release profiles of other experiments not used in the training. The results demonstrate that the Gaussian mixture model can be employed as a useful, intelligent tool for the prediction of time-series profiles in iontophoretic delivery systems.

## Summary and future directions

It is clear that the robust, nonlinear, model-free, universal approximator properties of neural networks make them very useful tools for modeling a wide range of systems relevant to drug development and delivery. The

applications of these methods in the study of ADMET processes and for building useful, predictive models will increase substantially in the future. The main limitations to more rapid progress in this area are not intrinsic to the method, but rather due to the lack of large, diverse data sets accessible to researchers developing new methods of modeling. The efficiency of current molecular descriptors and parsimonious methods for choosing the best set for a given modeling problem are also aspects to be improved. These areas are active research projects for most of the relatively small number of research groups who develop new modeling methodologies and models.

## References


1. Booth, B., Zemmel, R. *Prospects for productivity*. Nat Rev Drug Discov 2004, 3: 451-2.

2. DiMasi, J.A., Hansen, R.W., Grabowski, H.G. *The price of innovation: New estimates of drug development costs.* J Health Econom 2003, 22: 151-85.

3. Ekins, S., Boulanger, B., Swaan, P.W. et al. *Towards a new age of virtual ADME/TOX and multidimensional drug discovery*. J Comput-Aided Mol Des 2002, 16: 381-401.

4. Winkler, D.A. *The role of quantitative structure-activity relationships (QSAR) in biomolecular discovery*. Brief Bioinfo 2002, 3: 73-86.

5. Winkler, D.A., Madallena, D.J. *QSAR and neural networks in life sciences*. In: Series in Mathematical Biology and Medicine, Vol. 5. Computational Medicine, Public Health, and Biotechnology, Pt. 1. P.M. Auger, R.V. Jean (Eds.). World Scientific, Singapore, 1995, 126-63.

6. Winkler, D.A., Burden, F.R. *Application of neural networks to large dataset QSAR, virtual screening, and library design*. Meth Mol Biol 2002, 201: 325-67.

7. Matter, H., Baringhaus, K.-H., Naumann, T., Klabunde, T., Pirard, B. *Computational approaches towards the rational design of drug-like compound libraries*. Comb Chem High Through Screen 2001, 4: 453-75.

8. Zupan, J., Gasteiger, J., Meusinger, R. *Neural networks in chemistry and drug design.* Z Physik Chem 2000, 214: 1748-9.

9. Burden, F.R., Winkler, D.A. *New QSAR methods applied to structure-activity mapping and combinatorial chemistry*. J Chem Inf Comput Sci 1999, 39: 236-42.

10. Wrotnowski, C. *Counting on computational intelligence*. Mod Drug Discov 1999, 2: 46-8, 51-2, 55.

11. Bourquin, J., Schmidli, H., Van Hoogevest, P., Leuenberger, H. *Basic concepts of artificial neural networks (ANN) modeling in the application to pharmaceutical development*. Pharm Dev Technol 1997, 2: 95-109.

12. Agatonovic-Kustrin, S., Beresford, R. *Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research*. J Pharm Biomed Anal 2000, 22: 717-27.

13. Terfloth, L., Gasteiger, J. *Neural networks and genetic algorithms in drug design*. Drug Discov Today 2001, 6: S102-8.

14. Ichikawa, H. *Hierarchy neural networks as applied to pharmaceutical problems*. Adv Drug Deliv Rev 2003, 55: 1119-47.

15. Yamashita, F., Takayama, K. *Artificial neural network modeling for pharmaceutical research.* Adv Drug Deliv Rev 2003, 55: 1117.

16. Gasteiger, J., Teckentrup, A., Terfloth, L., Spycher, S. *Neural networks as data mining tools in drug design.* J Phys Org Chem 2003, 16: 232-45.

17. Polanski, J. *Self-organizing neural networks for pharmacophore mapping*. Adv Drug Deliv Rev 2003, 55: 1149-62.

18. Kirew, D.B., Chretien, J.R., Bernard, P., Ros, F. *Application of Kohonen neural networks in classification of biologically active compounds.* SAR QSAR Envir Res 1998, 8: 93-107.

19. Zupan, J. *2D mapping of large quantities of multi-variate data*. Croat Chem Acta 2002, 75: 503-15.

20. Bayada, D.M., Hamersma, H., van Geerestein, V.J. *Molecular diversity and representativity in chemical databases.* J Chem Inf Comput Sci 1999, 39: 1-10.

21. Bernard, P., Golbraikh, A., Kireev, D., Chretien, J.R., Rozhkova, N. *Comparison of chemical databases: Analysis of molecular diversity with self-organizing maps (SOM).* Analysis 1998, 26: 333-41.

22. Anzali, S., Gasteiger, J., Holzgrabe, U. et al. *The use of self-organizing neural networks in drug design.* Perspect Drug Discov Des 1998, 9/10/11: 273-99.

23. Manallack, D.T., Livingstone, D.J. *Neural networks in drug discovery: Have they lived up to their promise?* Eur J Med Chem 1999, 34: 195-208.

24. Livingstone, D.J., Manallack, D.T. *Statistics using neural networks: Chance effects.* J Med Chem 1993, 36: 1295-7.

25. Burden, F.R., Winkler, D.A. *Robust QSAR models using Bayesian regularized neural networks.* J Med Chem 1999, 42: 3183-7.

26. Burden, F.R., Winkler, D.A. *The computer simulation of high-throughput screening of bioactive molecules*. In: Molecular Modelling and Prediction of Bioactivity. K. Gundertofte, F.S. Jorgensen (Eds.). Plenum Press, New York, 2000.

27. Shah, A.V., Walters, W.P., Murcko, M.A. *Can we learn to distinguish between "drug-like" and "nondrug-like" molecules?* J Med Chem 1998, 41: 3314-24.

28. Bate, A., Lindquist, M., Edwards, I.R. et al. *A Bayesian neural network method for adverse drug reaction signal generation*. Eur J Clin Pharmacol 1998, 54: 315-21.

29. Topliss, J.G., Edwards, R.P. *Chance factors in studies of quantitative structure-activity relationships.* J Med Chem 1979, 22: 2.

30. Burden, F.R. *Using artificial neural networks to predict the biological activity of molecules from simple molecular structural considerations*. Quant Struct-Activ Relat 1996, 15: 7-11.

31. Winkler, D.A., Burden, F.R. *Robust QSAR models from novel descriptors and Bayesian regularized neural networks*. Mol Simul 2000, 24: 243-58.

32. Gasteiger, J. *Physicochemical effects in the representation of molecular structures for drug designing*. Mini-Rev Med Chem 2003, 3: 789-96.

33. Estrada, E., Uriarte, E. *Recent advances on the role of topological indices in drug discovery research.* Curr Med Chem 2001, 8: 1573-88.

34. Estrada, E., Patlewicz, G., Uriarte, E. *From molecular graphs to drugs. A review on the use of topological indices in drug design and discovery*. Ind J Chem 2003, 42A: 1315-29.

35. Huuskonen, J.J., Livingstone, D.J., Tetko, I.V. *Neural network modeling for estimation of partition coefficient based on atom-type electrotopological state indexes*. J Chem Inf Comput Sci 2000, 40: 947-55.

36. Hall, L.H., Kier, L.B. *Electrotopological state indices for atom types: A novel combination of electronic, topological and valence state information*. J Chem Inf Comput Sci 1995, 35: 1039-45.

37. Yasri, A., Hartsough, D.J. *Toward an optimal procedure for variable selection and QSAR model building.* Chem Inf Comput Sci 2001, 41: 1218-27.

38. Burden, F.R., Ford, M.G., Whitley, D.C., Winkler, D.A. *Use of automatic relevance determination in QSAR studies using Bayesian neural networks*. J Chem Inf Comput Sci 2000, 40: 1423-30.

39. Agrafiotis, D.K., Lobanov, V.S. *Nonlinear mapping networks*. J Chem Inf Comput Sci 2000, 40: 1356-62.

40. Lipinski, C.A., Lombardo, F., Dominy, B.W., Feeney, P.J. *Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings*. Adv Drug Deliv Rev 1997, 23: 3-25.

41. Walters, W.P., Murcko, M.A. *Prediction of "drug-likeness".* Adv Drug Deliv Rev 2002, 54: 255-71.

42. Murcia-Soler, M., Perez-Gimenez, F., Garcia-March, F.J., Salabert-Salvador, M.T., Diaz-Villanueva, W., Castro-Bleda, M.J. *Drugs and nondrugs: An effective discrimination with topological methods and artificial neural networks*. J Chem Inf Comput Sci 2003, 43: 1688-702.

43. Sadowski, J., Kubinyi, H. *A scoring scheme for discriminating between drugs and nondrugs*. J Med Chem 1998, 41: 3325-9.

44. Sadowski, J. *Optimization of the drug-likeness of chemical libraries*. Perspect Drug Discov Des 2000, 20: 17-28.

45. Sadowski, J. *Database profiling by neural networks*. Meth Prin Med Chem 2000, 10: 117-29.

46. Frimurer, T.M., Bywater, R., Nrum, L., Lauritsen, L., Norskov, B.S. *Improving the odds in discriminating "drug-like" from "non drug-like" compounds.* J Chem Inf Comput Sci 2000, 40: 1315-24.

47. Bruestle, M., Beck, B., Schindler, T., King, W., Mitchell, T., Clark, T. *Descriptors, physical properties, and drug-likeness*. J Med Chem 2002, 45: 3345-55.

48. Zernov, V.V., Balakin, K.V., Ivaschenko, A.A., Savchuk, N.P., Pletnev, I.V. *Drug discovery using support vector machines. The case studies of drug-likeness, agrochemical-likeness, and enzyme inhibition predictions*. Chem Inf Comput Sci 2003, 43: 2048-56.

49. Byvatov, E., Fechner, U., Sadowski, J., Schneider, G. *Comparison of support vector machine and artificial neural network systems for drug/nondrug classification*. J Chem Inf Comput Sci 2003, 43: 1882-9.

50. Takaoka, Y., Endo, Y., Yamanobe, S. et al. *Development of a method for evaluating drug-likeness and ease of synthesis using a data set in which compounds are assigned scores based on chemists' intuition*. J Chem Inf Comput Sci 2003, 43: 1269-75.

51. Taskinen, J., Yliruusi, J. *Prediction of physicochemical properties based on neural network modelling*. Adv Drug Deliv Rev 2003, 55: 1163-83.

52. Eros, D., Keri, G., Koevesdi, I., Szantai-Kis, C., Meszaros, G., Oerfi, L. *Comparison of predictive ability of water solubility QSPR models generated by MLR, PLS and ANN methods*. Mini-Rev Med Chem 2004, 4: 167-77.

53. Huuskonen, J., Salo, M., Taskinen, J. *Aqueous solubility prediction of drugs based on molecular topology and neural network modeling*. J Chem Inf Comput Sci 1998, 38: 450-6.

54. Ran, Y., Jain, N., Yalkowsky, S.H. *Prediction of aqueous solubility of organic compounds by the General Solubility Equation (GSE).* J Chem Inf Comput Sci 2001, 41: 1208-17.

55. Liu, R., So, S.-S. *Development of quantitative structure-property relationship models for early ADME evaluation in drug discovery. 1. Aqueous solubility*. J Chem Inf Comput Sci 2001, 41: 1633-9.

56. Manallack, D.T., Tehan, B.G., Gancia, E., Hudson, B.D., Ford, M.G., Livingstone, D.J., Whitley, D.C., Pitt, W.R. *A consensus neural network-based technique for discriminating soluble and poorly soluble compounds*. J Chem Inf Comput Sci 2003, 43: 674-9.

57. Bruneau, P. *Search for predictive generic model of aqueous solubility using Bayesian neural nets*. J Chem Inf Comput Sci 2001, 41: 1605-16.

58. DRAGON, Milano Chemometrics and QSAR Research Group. http://www.disat.unimib.it/chm/Dragon.htm.

59. Eros, D., Kovesdi, I., Orfi, L., Takacs-Novak, K., Acsady, G., Keri, G. *Reliability of logP predictions based on calculated molecular descriptors: A critical review*. Curr Med Chem 2002, 9: 1819-29.

60. Tetko, I.V. *Neural network studies. 4. Introduction to associative neural networks*. J Chem Inf Comput Sci 2002, 42: 717-28.

61. Devillers, J., Domine, D., Guillon, C. *Autocorrelation modeling of lipophilicity with a back-propagation neural network*. Eur J Med Chem 1998, 33: 659-64.

62. Huuskonen, J.J., Villa, A.E.P., Tetko, I.V. *Prediction of partition coefficient based on atom-type electrotopological state indices*. J Pharm Sci 1999, 88: 229-33.

63. Schaper, K.J., Rosado Samitier, M.L. *Calculation of octanol/water partition coefficients (logP) using artificial neural networks and connection matrixes*. Quant Struct-Act Relat 1997, 16: 224-30.

64. Doniger, S., Hofmann, T., Yeh, J. *Predicting CNS permeability of drug molecules: Comparison of neural network and support vector machine algorithms*. J Comp Biol 2002, 9: 849-64.

65. Fu, X.C., Wang, G.P., Liang, W.Q., Yu, Q.S. *Predicting blood-brain barrier penetration of drugs using an artificial neural network*. Pharmazie 2004, 59: 126-30.

66. Winkler D.A., Burden, F.R. *Modelling blood brain barrier partitioning using Bayesian neural nets*. J Mol Graph Mod 2004, 22: 499-508.

67. Podlogar, B.L., Muegge, I. *"Holistic" in silico methods to estimate the systemic and CNS bioavailabilities of potential chemotherapeutic agents*. Curr Top Med Chem 2001, 1: 257-75.

68. Agatonovic-Kustrin, S., Beresford, R., Yusof, A.P.M. *Theoretically-derived molecular descriptors important in human intestinal absorption*. J Pharm Biomed Anal 2001, 25: 227-37.

69. Wessel, M.D., Jurs, P.C., Tolan, J.W., Muskal, S.M. *Prediction of human intestinal absorption of drug compounds from molecular structure*. J Chem Inf Comput Sci 1998, 38: 726-35.

70. Niwa, T. *Using general regression and probabilistic neural networks to predict human intestinal absorption with topological descriptors derived from two-dimensional chemical structures*. J Chem Inf Comput Sci 2003, 43: 113-9.

71. Fujiwara, S.-I., Yamashita, F., Hashida, M. *Prediction of Caco-2 cell permeability using a combination of MO-calculation and neural network*. Int J Pharm 2002, 237: 95-105.

72. Gohlke, H., Dullweber, F., Kamm, W., Marz, J., Kissel, T., Klebe, G. *Prediction of human intestinal absorption using a combined "simulated annealing/back-propagation neural network" approach.* In: Rational Approaches to Drug Design. Proc 13th Eur Symp Quant Struct-Act Relat (Aug 27-Sept 1, Duesseldorf) 2000. H.-D. Hoeltje, W. Sippl (Eds.). Prous Science, Barcelona, 2001, 261-70.

73. Saxena, A.K., Schaper, K.-J. *QSAR analysis of the time dependent hypotensive activity of 1-[3-(4-substituted phenylthio)propyl]-4-(3,4-substituted phenyl)piperazines using artificial neural networks*. QSAR Comb Sci 2003, 22: 965-83.

74. Gobburu, J.V.S., Chen, E.P. *Artificial neural networks as a novel approach to integrated pharmacokinetic-pharmacodynamic analysis*. J Pharm Sci 1996, 85: 505-10.

75. Opara, J., Primozic, S., Cvelbar, P. *Prediction of pharmacokinetic parameters and their variability in bioavailability studies by artificial neural networks*. Farm Vest 1997, 48: 422-3.

76. Opara, J., Primozic, S., Cvelbar, P. *Prediction of pharmacokinetic parameters and the assessment of their variability in bioequivalence studies by artificial neural networks*. Pharm Res 1999, 16: 944-8.

77. Hussain, A.S. *Artificial neural network based in vitro-in vivo correlations*. Adv Exp Med Biol 1997, 423: 149-58.

78. Turner, J.V., Maddalena, D.J., Cutler, D.J. *Pharmacokinetic parameter prediction from drug structure using artificial neural networks*. Int J Pharm 2004, 270: 209-19.

79. Turner, J.V., Maddalena, D.J., Agatonovic-Kustrin, S. *Bioavailability prediction based on molecular structure for a diverse series of drugs.* Pharm Res 2004, 21: 68-82.

80. Veng-Pedersen, P., Modi, N.B. *Application of neural networks to pharmacodynamics.* J Pharm Sci 1993, 82: 918-26.

81. Haidar, S.H., Johnson, S.B., Fossler, M.J., Hussain, A.S. *Modeling the pharmacokinetics and pharmacodynamics of a unique oral hypoglycemic agent using neural networks.* Pharm Res 2002, 19: 87-91.

82. Ritschel, W.A., Akileswaran, R., Hussain, A.S. *Application of neural networks for the prediction of human pharmacokinetic parameters*. Meth Find Exp Clin Pharmacol 1995, 17: 629-43.

83. Moon, A., Smith, T. *A preliminary evaluation of neural network analysis for pharmacodynamic modeling of the dosing of the hydroxymethylglutaryl coenzyme A-reductase inhibitors simvastatin and atorvastatin*. Clin Ther 2002, 24: 653-61.

84. Wajima, T., Fukumura, K., Yano, Y., Oguma, T. *Prediction of human clearance from animal data and molecular structural parameters using multivariate regression analysis*. J Pharm Sci 2002, 91: 2489-99.

85. Schneider, G., Coassolo, P., Lave, T. *Combining in vitro and in vivo pharmacokinetic data for prediction of hepatic drug clearance in humans by artificial neural networks and multivariate statistical techniques*. J Med Chem 1999, 42: 5072-6.

86. Holmes, E., Nicholson, J.K., Tranter, G. *Metabonomic characterization of genetic variations in toxicological and metabolic responses using probabilistic neural networks*. Chem Res Toxicol 2001, 14: 182-91.

87. Vermeulen, N.P.E. *Prediction of drug metabolism: The case of cytochrome P450 2D6.* Curr Top Med Chem 2003, 3: 1227-39.

88. Zuegge, J., Fechner, U., Roche, O., Parrott, N.J., Engkvist, O., Schneider, G. *A fast virtual screening filter for cytochrome P450 3A4 inhibition liability of compound libraries.* Quant Struct-Act Relat 2002, 21: 249-56.

89. Zuegge, J., Schneider, G., Coassolo, P., Lave, T. *Prediction of hepatic metabolic clearance. Comparison and assessment of prediction models*. Clin Pharmacokin 2001, 40: 553-63.

90. Molnar, L., Keseru, G.M. *A neural network based virtual screening of cytochrome P450 3A4 inhibitors*. Bioorg Med Chem Lett 2002, 12: 419-21.

91. Korolev, D., Balakin, K.V., Nikolsky, Y. et al. *Modeling of human cytochrome P450-mediated drug metabolism using unsupervised machine learning approach*. J Med Chem 2003, 46: 3631-43.

92. Sorich, M.J., Miners, J.O., McKinnon, R.A., Winkler, D.A., Burden, F.R., Smith, P.A. *Comparison of linear and nonlinear classification algorithms for the prediction of drug and chemical metabolism by human UDP-glucuronosyltransferase isoforms*. J Chem Inf Comput Sci 2003, 43: 2019-24.

93. Sorich, M.J., McKinnon, R.A., Miners, J.O., Winkler, D.A., Smith, P.A. *Rapid prediction of chemical metabolism by human UPD-glucuronosyltransferase isoforms using quantum chemical descriptors derived with the electronegativity equalization method.* J Med Chem 2004, 47: 5311-7.

94. Schultz, T.W., Cronin, M.T.D., Netzeva, T.I. *The present status of QSAR in toxicology.* Theochem 2003, 622: 23-38.

95. Roche, O., Trube, G., Zuegge, J., Pflimlin, P., Alanine, A., Schneider, G. *A virtual screening method for prediction of the hERG potassium channel liability of compound libraries.* ChemBioChem 2002, 3: 455-9.

96. Feng, J., Lurati, L., Ouyang, H., Robinson, T., Wang, Y., Yuan, S., Young, S.S. *Predictive toxicology: Benchmarking molecular descriptors and statistical methods*. J Chem Inf Comput Sci 2003, 43: 1463-70.

97. Valkova, I., Vracko, M., Basak, S.C. *Modeling of structure-mutagenicity relationships: Counter propagation neural network approach using calculated structural descriptors.* Anal Chim Acta 2004, 509: 179-86.

98. Vracko, M., Mills, D., Basak, S.C. *Structure-mutagenicity modelling using counter propagation neural networks.* Environ Toxicol Pharmacol 2004, 16: 25-36.

99. Vracko, M., Szymoszek, A., Barbieri, P. *Structure-mutagenicity study of 12 trimethylimidazopyridine isomers using orbital energies and "spectrum-like representation" as descriptors.* J Chem Inf Comput Sci 2004, 44: 352-8.

100. Karelson, M., Sild, S., Maran, U. *Non-linear QSAR treatment of genotoxicity*. Mol Simul 2000, 24: 229-42.

101. Bahler, D., Stone, B., Wellington, C., Bristol, D.W. *Symbolic, neural, and Bayesian machine learning models for predicting carcinogenicity of chemical compounds.* J Chem Inf Comput Sci 2000, 40: 906-14.

102. Brinn, M., Walsh, P.T., Payne, M.P., Bott, B. *Neural network classification of mutagens using structural fragment data.* SAR QSAR Environ Res 1993, 1: 169-210.

103. Lindquist, M., Stahl, M., Bate, A., Edwards, I.R., Meyboom, R.H. *A retrospective evaluation of a data mining approach to aid finding new adverse drug reaction signals in the WHO international database.* Drug Saf 2000, 23: 533-42.

104. Agatonovic-Kustrin, S., Glass, B.D., Wisch, M.H., Alany, R.G. *Prediction of a stable microemulsion formulation for the oral delivery of a combination of antitubercular drugs using ANN methodology*. Pharm Res 2003, 20: 1760-5.

105. McCall, T., Chen, Y., Baichwal, A., Meyer, M. *Use of a neural network in CR product development.* Proc 24th Int Symp Control Rel Bioact Mat 1997, 267-8.

106. Sun, Y., Peng, Y., Chen, Y., Shukla, A.J. *Application of artificial neural networks in the design of controlled release drug delivery systems*. Adv Drug Deliv Rev 2003, 55: 1201-15.

107. Reis, M.A., Sinisterra, R.D., Belchior, J.C. *An alternative approach based on artificial neural networks to study controlled drug release*. J Pharm Sci 2004, 93: 418-30.

108. Lim, C.P., Quek, S.S., Peh, K.K. *Prediction of drug release profiles using an intelligent learning system: An experimental study in transdermal iontophoresis*. J Pharm Biomed Anal 2003, 31: 159-68.