Chapter 8
Neural Networks in Building QSAR Models

Igor I. Baskin, Vladimir A. Palyulin, and Nikolai S. Zefirov

Abstract This chapter critically reviews some of the important methods being used for building quantitative structure-activity relationship (QSAR) models using the artificial neural networks (ANNs). It attends predominantly to the use of multilayer ANNs in the regression analysis of structure-activity data. The highlighted topics cover the approximating ability of ANNs, the interpretability of the resulting models, the issues of generalization and memorization, the problems of overfitting and overtraining, the learning dynamics, regularization, and the use of neural network ensembles. The next part of the chapter focuses attention on the use of descriptors. It reviews different descriptor selection and preprocessing techniques; considers the use of the substituent, substructural, and superstructural descriptors in building common QSAR models; the use of molecular field descriptors in three-dimensional QSAR studies; along with the prospects of "direct" graph-based QSAR analysis. The chapter starts with a short historical survey of the main milestones in this area.

Keywords Artificial neural networks, QSAR, back-propagation, learning, generalization

1. Introduction

The first application of artificial neural networks (ANNs) in the domain of structure-activity relationships dates back to the early 1970s. In 1971, Hiller et al. [1] reported on a study dealing with the use of perceptrons, the only type of artificial neural networks known at that time [2], to classify substituted 1,3-dioxanes as active or inactive with regard to their physiological activity. In the cited work, coded elements of chemical structures were projected onto the perceptron retina; the perceptron was trained using a set of compounds with known activities, and the trained neural network demonstrated good recognition ability on both the training and the test sets of compounds. This methodology was discussed in detail in

D. S. Livingstone (ed.), *Artificial Neural Networks: Methods and Protocols*,

doi: 10/1007/978-1-60327-101-1 Springerprotocols.com

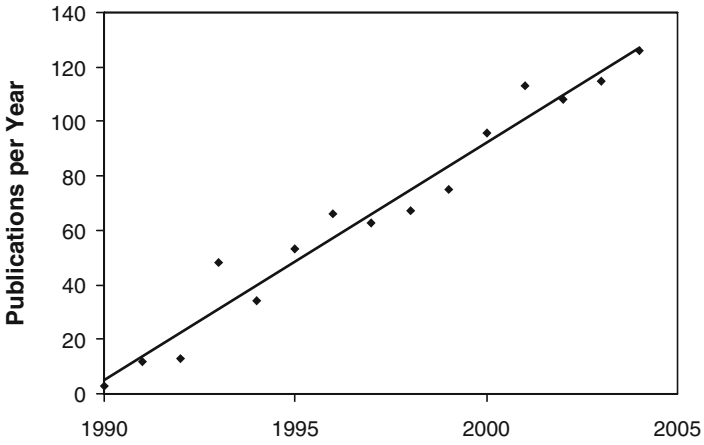

Fig. 8.1 Dependence of the number of papers published per year on the year of publication

another paper [3]. Nevertheless, the approach was not appreciated properly at that time; the articles remained almost unknown and have never been cited in any publication dealing with the use of neural networks in structure-activity studies.

The next stage of scientific development in this direction started in 1990 with the first publications of Aoyama, Suzuki, and Ichikawa dealing with the use of ANNs in QSAR studies [4, 5]. For the last 15 years, this approach to modeling structure-activity relationships has grown up and developed into a well-established scientific area with numerous ideas, theoretical approaches, and successful practical applications. Several relevant books [6, 7] and numerous review articles [8–28] are available. Figure 8.1 depicts the linear dependence of the number of papers published each year in this field on the year of publication. This linear trend demonstrates the steady growth of this scientific area, which now encompasses the use of artificial neural networks for predicting not only different types of biological activity of chemical compounds but also their physicochemical, ADME, biodegradability, and spectroscopic properties, as well as their reactivity. The aim of this paper is to review some important concepts and ideas accumulated in this field.

2. Methods, Discussion, and Notes

2.1. Neural Network Computational Methods Used in QSAR Studies

Applications of artificial neural networks in QSAR studies are characterized by a wide variety of architectures of neural networks as well as numerous approaches to represent chemical structures, preprocessing and selection of relevant descriptors,

running the learning process, handling the predicted results, and so forth. All these notions characterize the *computational methods* used in this scientific area. The notion of the computational method is more specific than the notions of the neural network *architecture* and neural network *paradigm*. For example, we treat the genetic, the Bayesian, and the ordinary back-propagation neural networks as distinct computational methods, although they actually use the same architecture. Table 8.1 lists the main computational neural network methods with at least one application in QSAR studies. For each method, the table shows its name, a short identifier (used further in this chapter), the year of the first publication dealing with its use in QSAR studies, as well as the reference to such publication.

Table 8.1 Neural network methods used in QSAR studies

Name	Identifier	Year	Reference
Perceptron	Perceptron	1971	[*1*, *3*]
Back-propagation neural network	BPNN	1990	[*4*]
Autoassociative feedforward neural network	AAFFNN	1991	[*29*]
Self-organizing maps (Kohonen neural network)	SOM	1991	[*30*]
Counterpropagation neural network	CPNN	1992	[*31*]
Function-link neural network	FUNCLINK	1992	[*32*]
Neural device for direct structure-property correlation	NDDSPC	1993	[*33*, *34*]
Radial basis functions neural network	RBFNN	1993	[*35*]
Evolutionary algorithm for BPNN	Ev-BPNN	1993	[*36*]
Ensembles of BPNNs	Ens-BPNN	1993	[*37*]
Simulated annealing for BPNN	SA-BPNN	1995	[*38*]
Genetic algorithm for BPNN	GA-BPNN	1996	[*39*]
Principal component analysis for BPNN	PCA-BPNN	1996	[*40*]
Adaptive resonance theory 2-A	ART-2-A	1997	[*41*]
Bayesian regularized neural network	BRNN	1997	[*42*]
Probabilistic neural network	PNN	1997	[*43*]
Receptor-like neural network	RLNN	1997	[*44*]
Cascade-correlation neural network	CCNN	1998	[*45*]
Genetic algorithm for CPNNs	GA-CPNN	1999	[*46*]
Fuzzy adaptive resonance theory for mapping	FARTMAP	2000	[*47*]
Fuzzy neural network	FNN	2000	[*48*]
Recursive cascade correlation neural network	RCCNN	2001	[*49*]
Volume learning algorithm neural network	VLANN	2001	[*50*]
Comparative molecular surface analysis	CoMSA	2002	[*51*]
Genetic algorithm for RBFNN	GA-RBFNN	2002	[*52*]
Generalized regression neural network	GRNN	2002	[*53*]
Integrated SOM-fuzzy ARTMAP neural system	SOM-FARTMAP	2002	[*54*]
Particle swarms for BPNN	PS-BPNN	2002	[*55*]
Artificial ant colonies for BPNN	ANT-BPNN	2002	[*56*]
Hopfield neural network	HNN	2003	[*57*]
Genetic algorithm and principal component analysis for back-propagation neural network	PCA-GA-BPNN	2003	[*58*]
Variable structure neural net for pattern recognition	VSNNPR	2003	[*59*]
Niching particle swarms for BPNN	NPS-BPNN	2003	[*60*]
Genetic algorithm for self-organizing maps	GA-SOM	2004	[*61*]
Learning vector quantization	LVQ	2004	[*62*]

2.2. Tasks Performed by Neural Networks in QSAR Studies

In principle, ANNs can be used for solving any solvable task in computational mathematics, ranging from simple addition of binary numbers up to theorem proving. A special field in the computer science, *neuromathematics*, tackles such problems. In practice, ANNs are usually used for solving so-called ill-posed problems, for which numerous alternative solutions can be suggested, such as function approximation, pattern recognition, clustering, and data reduction. These problems exactly correspond to tasks performed by neural networks in an absolute majority of QSAR studies. Solving typical ill-posed problems involves some sort of *learning* or *training*, which can be *supervised*, *unsupervised*, or *reinforcement*. The last type of learning has not yet been used in QSAR/QSPR applications. Unsupervised learning analyzes the internal data structure by finding clusters and reducing dimensionality. In this paper, we discuss only supervised learning.

2.3. Supervised Learning

In the course of the supervised learning, a neural network tries to approximate experimental data by comparing its output values with the "desired" ones and minimizing the discrepancy by adjusting its internal parameters, such as connection weights and activation thresholds. If the experimental data are expressed by real numbers, the network performs *function approximation*, while for discrete, especially binary, values it performs *pattern recognition*. So, in statistical terms, it performs *regression* analysis and data *classification* (discriminant analysis), respectively. QSAR studies usually deal with regression analysis, while data classification is carried out in SAR studies. The rest of this chapter attends to the regression analysis performed by ANNs in QSAR studies.

2.3.1. Regression

The regression task is solved in almost all quantitative structure-property relationships (QSPR) and in a big part of structure-activity applications. The most widely used computational method for this purpose is BPNN (feedforward neural net trained with error back propagation, or back-propagation neural networks) or one of its derivatives (such as BRNN or GA-BPNN). The popularity of BPNN originates from its ability to be a "model-free mapping device" [63], which can approximate any nonlinear dependence of output variables (properties, activities) on the values of input variables (usually descriptors). Kolmogorov's superposition theorem [64] on the representation of continuous functions of many variables by superposition of continuous functions of one variable and addition, in conjunction with Kurková's results [65], presents the basis for this universal approximating ability of BPNNs.

So, the main advantage of using multilayer ANNs in QSAR studies lies in their ability to cope with nonlinearity in relationships, this ability requiring a hidden layer of neurons. Neural nets without hidden neurons, such as FUNCLINK, can hardly be considered as universal approximators, although some kinds of nonlinearity can be revealed by them due to nonlinear transforms of input variables (see the discussion in [66]). BPNNs approximate properties *globally*, since all hidden neurons are involved in production of neural-net output signal for all input vectors. In contrast, some other kinds of ANNs, such as CPNN or RBFNN, approximate properties *locally*, since only few neighboring (in descriptor space) hidden neurons are really involved in production of neural-net output. As a result, in spite of the ability of CPNNs and RBFNNs to approximate any nonlinear function with any precision (as a consequence of Kolmogorov's theorem), the number of hidden neurons required for attaining a given approximation error depends, in the general case, exponentially on data dimensionality. This is a consequence of the *dimensionality curse* [67], which is inherent to nonparametric kernel-based approximators. Since the number of hidden neurons should be much smaller than the number of data points, which is always very limited in QSAR studies, the approximating ability of BPNNs outperforms that of CPNNs and RBFNNs for data sets with not very low data dimensionality.

The global approximation character of BPNNs has some additional important implications for QSAR studies. If the compounds used for training and predictions are rather distinct from each other, one cannot expect good predictive performance of any statistical or neural-net model. However, if the compounds from the prediction set are similar to each other and the property to be predicted is measured for some of them, then it is possible to correct predictions for the remaining compounds from the prediction set by utilizing the measured property values of similar compounds without the need to retrain the neural network. This idea lies behind the associative neural networks (ASNN) concept [68], which combines global approximation performed by BPNNs with a local nearest-neighbors correction. In ASNN, the similarity of chemical structures is computed in the model space by correlating predictions made by different BPNN models. We implemented a different procedure in the NASAWIN program, which consists in correcting BPNN predictions by utilizing experimental property values of neighboring compounds, which maps onto the same cell in Kohonen neural network (SOM) trained using the same or another set of descriptors [59]. Raevsky et al. suggested correcting MLR (multiple linear regression, which performs global approximation) predictions with nearest-neighbor corrections, the similarity being determined by means of Tanimoto coefficients computed using substructure screens [69]. So, local correction to global structure-property approximations can be very useful, and it can efficiently be utilized by BPNNs.

The global approximating character of BPNNs has a direct connection to the distributed (holographic) character of information storage in human brain. As a consequence, the influence of an input signal on an output one is coded by means of several connection weights, and therefore the values of individual connection weights, taken separately, tell nothing about the influence of any input signal on any

output one. In other words, weights provide nonlocal information on the influence of an input variable [70]. Consequently, individual connection weights cannot be interpreted separately from the other ones, and this has given rise to the myth of uninterpretability of neural network models.

2.3.2. Interpretation of Neural Network Regression Models

The problem of neural network *interpretation* has been addressed in a number of studies (see [70, 71] for relevant references). Aoyama and Ichikawa suggested using the partial differential coefficients of output value with respect to input parameters to analyze the relationship between inputs and outputs in neural network for each data point and demonstrated this method for QSAR data on biological activity of mitomycins [72]. Developing these ideas further, we suggested in a paper [73] to use the first (mean value) and the second (dispersion) distribution moments of the first and the second partial derivatives of the outputs with respect to inputs in neural networks over the whole training set for interpreting neural network QSAR/QSPR models. The use of such statistics makes it possible not only to obtain actually the same characteristics as for traditional "interpretable" statistical methods, such as the linear regression analysis, but also to reveal important additional information concerning the nonlinearity of QSAR/QSPR relationships. This approach was illustrated by interpreting BPNN models for predicting position of the long-wave absorption band of cyan dyes [73] and the acid hydrolysis rate constants of esters [74]. In both cases, the interpretations of neural network models appeared to be compatible with theoretical knowledge about the underlying physical and chemical phenomena.

Guha et al. have recently readdressed this problem in two papers [71, 75]. Guha, and Jurs [75] presented a method to measure the relative importance of the descriptors in QSAR neural network models based on a sensitivity analysis of descriptors. For all the reported data sets, the important descriptors revealed by this method appeared to correspond to the important descriptors in the linear model built using the partial least squares (PLS) technique. This means that the interpretability of neural network models is not reduced in comparison with traditional statistical linear approaches (exemplified by the PLS method), at least for interpretation methods based on ranking the relative importance of descriptors. Guha, Stanton, and Jurs [71] presented a methodology to carry out a detailed interpretation of the weights and biases of neural network QSAR models. It allows one to understand how an input descriptor is correlated with the predicted output property. The proposed methodology is based on the linearization of the transform functions in all computational (hidden or output) neurons. The data flow in the resulting network mimics that of the PLS analysis, with the latent variables of the latter corresponding to the hidden neurons in the network and the appropriate matrix elements corresponding to the connection weights in the neural network. This allowed the authors to develop interpretations of a neural

network regression model similar in manner to the partial least squares interpretation method for linear models described by Stanton [76]. The method was tested on several data sets, the results of the interpretation method corresponded well to PLS interpretations for linear models using the same descriptors, and they were shown to be consistent with the generally accepted physical interpretations for these properties.

So, artificial neural networks (at least BPNNs) should no longer be regarded as uninterpretable "black boxes," and each QSAR work involving development of neural network models would benefit from application of interpretation procedures.

2.3.3. Methods for Controlling Generalization

Generalization means the ability of neural networks to predict the observed response variable for patterns not included in the training set. In other words, this is the ability to predict some biological activity or some other property of new chemical compounds. *Generalization error* can be estimated by computing the root-mean-square error of prediction on some external *validation* set. *Memorization* means the ability to reproduce the values of the response variable for patterns taken from the training set. *Memorization error* is the root-mean-square error of prediction on the training set. Evidently, in QSAR studies the primary concern should be to build neural network models that generalize better, that is, with the smallest generalization error. If, for some fixed data set, we gradually increase the complexity of some neural network (which is defined as the number of its adjustable parameters, i.e., connection weights and biases) by adding additional hidden neurons, the generalization error initially decreases but, after reaching the optimal network size, starts to increase, although the memorization error decreases all the time. The resulting neural network has bad generalization and good memorization ability. This phenomenon is called *overfitting*. Generally, the training of a neural network is a multistep iterative process. At the first phase of the training, the generalization error decreases but after reaching some point starts to increase, although the memorization error decreases all the time. Again, the resulting neural network model shows bad generalization but good memorization ability. This phenomenon is called *overtraining*. Therefore, one should exert every effort to prevent overfitting and overtraining. Both phenomena are thoroughly analyzed in papers [77–80] as applied to building QSAR models. The recipe to prevent overfitting, in accordance to these papers, lies in keeping the values of the parameter ρ, which is the ratio of the number of data points to the number of connections, higher than some threshold. This parameter, put forward originally by Andrea and Kalayeh [81], was analyzed in detail by Livingstone and coworkers [77, 78]. The overtraining can be avoided by means of "early stopping" of training after reaching the lowest generalization error as estimated using an additional validation data set [80]. It is also asserted by Tetko, Livingstone, and Luik [80] that overfitting and the associated "chance effects" are also prevented with the prevention of overtraining.

2.3.4. The Minimum Description Length Principle

To understand the essence of the aforementioned phenomena and assess the aforementioned results, consider the neural network learning from the statistical point of view [82]. According to the Bayesian theorem,

$$P(N \mid D) = \frac{P(D \mid N)P(N)}{\sum_N P(D \mid N)P(N)} \tag{1}$$

where N is a neural network model, D is a data set used for its training, $P(N \mid D)$ is the probability (i.e., validity) of the neural network model N trained with data D, $P(D \mid N)$ is the probability of data D to be explained by the neural network model N, and $P(N)$ is an a priori probability of the neural network model N. The best neural network model can be found by maximizing $P(N \mid D)$ or its logarithm:

$$\max_N \log P(N \mid D) \Rightarrow \max_N \{\log P(D \mid N) + \log P(N)\} \tag{2}$$

This is equivalent to

$$\min_N \{-\log P(D \mid N) - \log P(N)\} = \min_N \{I_D + I_N\} \tag{3}$$

where I_D is the memorization error, expressed as the number of information bits necessary to account for all residuals on the training set, while I_N is the *model complexity*, expressed as the quantity of information needed for choosing the model N from the set with the a priori probability distribution $P(N)$. For multilayer neural networks, I_N can be taken as proportional to the number of connections and therefore to the number of hidden neurons. Equation 3 forms the basis of the minimum description length (MDL) principle originally put forward by Rissanen [83]. According to this principle, models with the shortest joint description of data and the model built on these data provide the lowest generalization error. Such models are characterized by the lowest error on the training set and the minimal possible number of adjustable model parameters.

The essence of the neural network training lies in the minimization of I_D. After the training, I_D becomes comparable or even lower than I_M. So, if the number of hidden neurons in BPNN exceeds some optimum, I_M starts to dominate over I_D, and the generalization error tends to be high. This explains the overfitting phenomenon and also opens the ways to finding the optimal number of connections W and hidden neurons H. It can be shown that for a simple case W can be estimated as [82]

$$W \sim \sqrt{Pd} \tag{4}$$

where P is the number of points in the training set, d is the number of input variables. More correct treatment would also include the dependence of W on some additional parameters, such as the dimensionality of d and the complexity of

a function being approximated. In any case, it is not sufficient to use the simple ρ parameter as a criterion for the overfitting.

The MDL principle can also be used for explaining the overtraining phenomenon. The neural network training usually starts with all adjustable parameters initialized with random values close to zero. The complexity I_N of such model is also close to zero. In the course of training, I_N gradually increases. At the initial phase of the training, when the values of all connection weights are still small, the responses of all transfer functions in neurons lie in the linear range, and therefore the network approximates the linear part of relationships. The effective number of adjustable parameters at this linear phase does not exceeds that of a fully trained linear model, that is, approximately *d*. So, at the first phase, the effective number of adjustable parameters *gradually* increases from 0 to approximately *d*. After that, with the rise of the connection weight values, the responses of neurons become more and more nonlinear, the network starts learning nonlinearity, and the effective number of adjustable parameters *gradually* increases from *d* to approximately $H \cdot d$. At some moment, I_N starts to rise quicker than I_D decreases in the course of the training. This explains the overtraining phenomenon.

Several conclusions can be drawn concerning the overtraining phenomenon. First, although "early stopping" is an important procedure, which is usually necessary for preventing overtraining and overfitting, its application results in models with reduced nonlinearity. The bigger is the size of the network, the more linear are such models. Therefore, the "early stopping" should always be applied with caution. This "overlinearization" phenomenon can be prevented by applying the automatic weight elimination (AWE) algorithm [82]. Second, application of very fast algorithms for neural network training is not always the best choice, since they may quickly produce overtrained models without the possibility to avoid this.

2.3.5. Regularization of ANN Models

An alternative approach to preventing overtraining lies in the use of *regularization*, which is introduced by means of special functions that penalize the growth of the connection weights and, by doing that, limit the growth of the model complexity I_N. This results in the decrease of the generalization error. The regularization method initially was developed by Tikhonov and Arsenin as a general principle for solving ill-posed problems [84]. In this method, the influence of the penalty function is controlled by a special regularization parameter, the optimal value of which can be found either by assessing generalization using cross-validation or by means of the Bayesian approach through statistical analysis of connection weight distributions without the need to use cross-validation. So, the latter approach, called *Bayesian regularization*, appeared to be very useful for building QSAR models (see, for example, [85]), since it uses the whole data set for training and the resulting neural network models are reproducible. Two other advantages of Bayesian neural networks are the possibility to apply the procedure of automatic relevance determination for selecting descriptors and the possibility to use the distributions of the predictions to evaluate their uncertainty.

2.3.6. Ensembles of Neural Networks

One of the drawbacks of using BPNN in QSAR studies lies in irreproducibility of the resulting QSAR models. Starting with the initial random values of connection weights, they can produce different models. This problem can be solved by using an ensemble of neural networks models instead of a single one and averaging the prediction results. As a result, the models become more reproducible, the chance effects caused by initial randomization of weights are eliminated, and the memorization and generalization errors usually become lower. This methodology was initially applied to building QSAR models in 1993 by Tetko, Luik, and Poda [37]. The issue of using of the neural network ensembles in QSAR and QSPR studies was analyzed in detail by Agrafiotis, Cedeño, and Lobanov [86].

2.4. Handling Chemical Structures

2.4.1. Descriptor Selection

Because of the abundance of different descriptor selection approaches used in QSAR studies in conjunction with ANNs, we list them while applying some sort of classification. Depending on the use of activity values, descriptor selection can be unsupervised or supervised. The *unsupervised forward selection* procedure, which generates a subset of nonredundant descriptors with reduced multicollinearity, is described by Whitley, Ford, and Livingstone [87]. A sort of an *unsupervised backward elimination* procedure, which, starting with the whole descriptor set, leaves a subset of descriptors with pairwise correlations limited by some threshold value (although with multicollinearity) is implemented in the NASAWIN software [59].

Supervised descriptor selection procedures can be preliminary (carried out before running neural networks) or online. A *supervised preliminary forward selection* procedure based on the use of stepwise linear regression analysis procedures is implemented in NASAWIN [59]. It works as follows. The first chosen descriptor provides the best correlation with the target property, while each next selected descriptor provides the best correlation with the residual vector obtained at the previous step. The procedure is stopped after a fixed number of steps or at the lowest prediction error for the validation set. The main advantage of the algorithm lies in its very high efficiency (it can easily operate with millions of descriptors on very large databases), although the quality of the resulting models is slightly worse in comparison with the standard stepwise multiple linear regression technique. Nonetheless, the resulting suboptimal subset of descriptors appears to be sufficiently good as an initial guess for conducting not only regression but also classification SAR/QSAR studies using BPNNs.

Supervised online descriptor selection procedures, which are executed simultaneously with the ANN learning, can be NN-guided or controlling. The *supervised online NN-guided forward selection* procedures are performed by means of

"growing" neural network architectures, such as cascade-correlation networks (CCNNs) for approximation [45]. The *supervised online NN-guided backward elimination* procedures are based on the use of various *pruning* algorithms for eliminating unnecessary weights and neurons from neural networks (see, for example, [88]).

The *supervised online controlling descriptor selection* procedures are usually implemented by means of stochastic combinatorial optimization algorithms, which optimize some measure of neural network model quality by finding optimal subsets of selected descriptors. Probably the first application of this kind of descriptor selection procedures in SAR/QSAR studies was made by Brinn et al. in 1993 using the "evolution algorithm" (Ev-BPNN) [36]. This was followed by the application of the simulated annealing (SA) algorithm for the same purpose [38]. The next important step was made in 1996 by So and Karplus [39] by introduction of the *genetic algorithm* (GA), which, being combined with different neural network architectures, has led to numerous computational methods actively used in QSAR studies: GA-BPNN ("genetic neural network") [39], GA-CPNN [46], GA-RBFNN [52], GA-SOM [61]. Several modern combinatorial optimization algorithms have also been applied in combination with BPNNs for descriptor selection in QSAR studies: the *artificial ant colonies* (ANT-BPNN) [56], the *particle swarms* (PS-BPNN) [55], and the *niching particle swarms* (NPS-BPNN) [60]. So, a big arsenal of methods is currently available for performing descriptor selection in QSAR studies.

2.4.2. Descriptor Preprocessing

Three principal types of data preprocessing are used in QSAR studies: scaling, nonlinear transform, and reduction of data dimensionality. While the scaling of input and output variables is a standard practice and does not deserve special consideration in this paper, the issue of the nonlinear transform is not trivial. Although one might argue that nonlinear transfer is not needed with neural networks, nevertheless our practice is the evidence of its benefit [89]. Two plausible explanations for this can be given. The first deals with the use of the early stopping procedure for preventing overtraining, which hampers the learning of strong nonlinearity, while the second explanation concerns our use of the preliminary descriptor selection procedure based on the use of stepwise linear regressions, which can partially account for nonlinearity due to nonlinearly transformed descriptors (like in the function-link neural networks).

Data dimensionality reduction is usually carried out in QSAR studies by means of principal component analysis (PCA). This gave rise to the PCA-BPNN [40] and PCA-GA-BPNN [58] computational methods, although combination with other neural network architectures is also possible. SOMs are also used for this purpose, like in SOM-FARTMAP [54]. We successfully used PLS latent variables for input into BPNNs [59]. However, several promising modern data preprocessing techniques, such as the independent component analysis (ICA) [90], are still not reported to have been used in QSAR studies in conjunction with ANNs.

2.4.3. Substituent Descriptors

A big part of QSAR studies deals with the analysis of congeneric sets of compounds with common scaffolds and different substituents. The classical Hansch-Fujita [91] and Free-Wilson [92] approaches imply the use of substituent constants, such as Hammett σ constants and lipophilic constants π, as well as variables indicating the presence of some structural features at some fixed scaffold positions, for deriving QSAR models. As applied to these substituent-based approaches, artificial neural networks have efficiently been used, in addition to building QSAR models, for solving two problems specific for this kind of descriptor: prediction of substituent constants and construction of QSAR models with correct symmetry properties.

2.4.3.1. Prediction of Substituent Constants

Conducting QSAR studies in the framework of the Hansch-Fujuta approach involves the use of substituent constants, which are tabulated for many but not all possible substituents. This limits the practical use of this approach to derivatives with simple substituents and issues the challenge of predicting the values of various substituent constants for any substituent. To address this problem, Kvasnička, Sklwnák, Pospichal applied BPNNs trained with specially designed functional group descriptors (occurrence numbers of some labeled subgraphs) to predict the values of inductive (σ_I) and resonance (σ_R) substituent constants [93]. A training set consisting of 66 substituents with tabulated values of both constants was used in their study. Several years later, we used BPNNs and quantum-chemical descriptors to build neural network models for predicting five substituent constants: σ_m, σ_p, F, R, and E_s [94]. Two BPNNs were used in this study. The first one, with four outputs for simultaneous prediction of σ_m, σ_p, F, and R, was trained using data on 144 substituents, while the second one, with a single output and trained with data on 42 substituents, was used for predicting E_s. Good predictive performance was achieved for all cases.

The problem was recently readdressed by Chiu and So [95, 96]. They used a big data set of 764 substituents for training BPNNs to predict four Hansch substituent constants: π, MR, F, and R [95]. The E-state descriptors were used for π and MR, while the aforementioned Kvasnička's graph-theoretical functional group descriptors were used for F and R. In the subsequent paper [96], the values of substituent constants predicted using these neural network models were successfully used for deriving QSAR models for HIV-1 reverse transcriptase and dihydrofolate reductase inhibitors. So, the subsequent use of ANNs for computing descriptors and conducting QSAR studies enabled the authors to obtain easily interpretable QSAR models with good predictive performance [96]. This pair of works may constitute the first example of the hierarchical approach to QSAR modeling, which seems to be very promising one.

2.4.3.2. Building QSAR Models with Correct Symmetry Properties

The next problem associated with the use of substituent descriptors lies in the necessity to construct QSAR models with correct symmetry properties. This means that, if the scaffold of some congeneric series of compounds is symmetric and contains topologically equivalent substituent positions, then the predicted values of any property, including biological activities, should not depend on the way nonequivalent substituents are assigned to them. For example, if positions 1 and 2 in some scaffold are topologically equivalent, then the assignments (R_1 = Cl, R_2 = Br) and (R_1 = Br and R_2 = Cl) designate the same compound, and QSAR models with correct symmetry properties should predict for them the same activity values.

This issue was analyzed by us [97]. To tackle the problem, we put forward an approach based on the application of ANNs to the training sets expanded by adding the copies of compounds with the same activity values but with permuted assignment of equivalent substituent positions (the learned symmetry concept). As the proof of the concept, the better predictive ability of resulting QSAR models, as compared with the performances of neural network models for nonexpanded sets, was demonstrated for the calcium channel blockers of 1,4-dihydropyridine type and hallucinogenic phenylalkylamines.

2.4.4. Substructural Descriptors

Among all other kinds of descriptors used in conjunction with ANNs, the substructural ones (i.e., the occurrence numbers of some subgraphs in molecular graphs) occupy a special place. As we proved earlier [98, 99], any molecular graph invariant (that is, any nonlocal property of a chemical compound) can be uniquely represented as (1) a linear combination of occurrence numbers of some substructures (fragments), both connected and disconnected, or (2) a polynomial on occurrence numbers of some connected substructures. Since, according to Kolmogorov's theorem [64], any polynomial can be approximated with a three-layer neural network, any property that is not very sensitive to stereoisomerism (such as a majority of physical properties) can be approximated by an output of a multilayer, globally approximating neural network taking occurrence numbers of connected substructures as its inputs. We used this as a guideline in our studies on predicting various physicochemical properties of organic compounds [89, 100–105].

The main shortcoming of the direct use of substructural descriptors in QSAR studies lies in the necessity to possess a sufficiently large database, in which all important fragments should be well-represented. That is why substructural descriptors are used predominantly for predicting properties, for which such databases are available, namely, the physicochemical, and toxicological ones. For an example of the use of fragmental descriptors in conjunction with multilayer neural networks and a large structurally heterogeneous data set for predicting mutagenicity, see [36]. On the other hand, if only small data sets are available, substructural

descriptors can be very effective in mixture with physicochemical and quantum-chemical ones (for an example, see our work on mutagenicity of substituted polycyclic compounds [106]).

2.4.5. Superstructural Descriptors

Superstructural descriptors constitute a good alternative to substructural ones for building QSAR models for small data sets. They are computed by mapping molecular graphs into some molecular supergraph (for the whole data set), transferring their local atomic and bond descriptors to the corresponding supergraph elements, from which fixed-sized descriptor vectors are formed. This methodology can efficiently be combined with the use of ANNs, as was demonstrated in [107] for the cases of building neural network QSAR models for phospholipase inhibition by substituted indoles, dopamine receptor antagonistic activity of quinolinones, anti-HIV activity of benzylpyrimidines, and the ability of peptidyl trifluoromethyl ketones to prevent elastase-induced lung damage.

2.4.6. Molecular Field Descriptors for Three-Dimensional QSAR

Although computation of a big part of different molecular descriptors uses information on molecular geometry, the term *3D-QSAR* is traditionally associated with CoMFA and molecular field descriptors (i.e., the values of electrostatic, steric, lipophilic, and similar potentials computed at some points in space, e.g., at lattice nodes or on a molecular surface) involved in it [108]. We confine our discussion to the use of this kind of descriptors. Several approaches to using ANNs in 3D-QSAR studies have been developed (see the review in [21]). The main milestones in this field are Polanski's receptor-like neural network (RLNN) [44], which utilizes charges located on atoms as space-positioned descriptors; the volume learning algorithm neural network (VLANN) developed by Tetko, Kovalishyn, and Livingstone [50], which uses molecular field descriptors computed as some points in space; the comparative molecular surface analysis (CoMSA) put forward by Polanski, Gieleciak, and Bak [51], which uses molecular field descriptors computed on a molecular surface; and a molecular surface-based method developed by Hasegawa and coworkers [109], which is similar to CoMSA and also uses molecular surface descriptors. Despite all the differences among these four methods, they adopt actually the same basic two-stage strategy with molecular field potentials being mapped from their initial points in Cartesian space onto the cells of the Kohonen neural network at the first step, followed by the application of some regression technique, such as the committee of BPNNs in VLANN and PLS in the other three procedures, to correlate the potential values averaged over the cells with the target property at the second stage of this methodology. Consider, however, the mentioned approaches from the view point of data processing.

The specificity of the 3D-QSAR studies lies in the large number (typically, thousands) of spatially organized molecular field descriptors. So, the main two challenges in using ANNs are to sharply reduce the data dimensionality and to introduce some sort of nonlinearity into QSARs [21]. While the second task is tackled in VLANN by means of BPNNs, the first issue is addressed in all three approaches by means of using the Kohonen self-orginizing maps. One might suppose that it is the unique property of SOMs to reduce drastically the data dimensionality used in these applications. Surprisingly, but this appears not to be the case. The number of variables is actually reduced in these methods by means of their clustering with the help of SOMs. Furthermore, such reduction is performed inefficiently in comparison with the well-established CoMFA approach, since the number of cells in SOMs is always much greater than the number of latent variables in CoMFA PLS analysis. SOMs are known to be very effective tools for performing topology analysis of data sets and their neighborhood-conserving mapping onto the graphs of interconnected neurons in the competitive layer. The success in describing data by means of SOMs greatly depends on the adequacy of topologies embedded in the SOMs to that of the data being analyzed. This immediately poses the question as to whether the toroidal topology with four neighbors for each neuron, which is currently adopted in all SOM applications in SAR/QSAR studies, is adequate to represent the geometry of molecules and the topology of molecular fields around them. Would not the spherical topology of neurons be more adequate for self-organizing mapping of molecular fields? Is it so necessary to confine such mapping to any topology at all? The main shortcoming of using SOMs for representing molecular fields lies in the abstract nonphysical characters of such maps, which can hardly be understood by chemists and biologists outside the QSAR community. And, finally, is competitive learning needed at all to preprocess data for 3D-QSAR analysis? Our preliminary computational experiments indicate the feasibility of various alternative approaches to using ANNs in 3D-QSAR studies [110].

So, in spite of the existence of a whole series of remarkable applications of ANNs in the area of 3D-QSAR analysis, this field is still in its infancy. It waits for new ideas to be expressed, new methods to be developed, and new striking applications to be contributed.

2.4.7. Vector-Based and Graph-Based QSAR

All QSAR approaches considered in this paper so far are vector-based, since the descriptors in them should be represented as vectors of the same size for each of the chemical compounds belonging to the same data set. Almost all statistical procedures and ANNs were vector-based till recently. However, the most natural mathematical objects for representing the structures of chemical compounds are molecular graphs or the corresponding matrices of *variable size*. Since all isomorphic molecular graphs correspond to the same chemical compound, any structure-activity relationship function *should not depend on the numbering of nodes in molecular graphs* (or the permutation of rows and columns in the corresponding

connection tables). A traditional approach to building QSAR models consists in computing vectors of molecular descriptors (graph *invariants*, whose values do not depend on the numbering of nodes in molecular graphs), which are usually chosen ad hoc, followed by the application of a vector-based statistical or ANN technique for finding QSAR models. As an obvious shortcoming of the traditional approach, the resulting models appear to be too biased and too dependent on the choice of a necessary descriptor set.

As an alternative to the traditional vector-based QSAR approach, an interesting challenge would be to build graph- or matrix-based QSAR models. To address this problem, in 1993, we developed a special neural device (NDDSPC) capable of constructing graph-based QSPR relationships for alkanes [33]. An advanced version of NDDSPC was used further for constructing a number of graph-based QSPRs and QSARs for heterogeneous sets of chemical compounds [34]. In 1995, Kireev proposed the graph-based ChemNet neural network, capable of establishing QSPRs [111]. The next contribution was made by Ivanciuc with his MolNet, which can also be used for building QSPR models [112].

By the end of the 1990s, the necessity of creating machine learning methods capable of handling variable-size structured data, such as chemical structures or biological sequences, was realized by computer scientists, and this led to important developments in this field (the eighth issue of the *Neural Networks* journal in 2005 completely deals with applications in this currently very hot scientific area). Two types of neural networks, probabilistic Bayesian networks and deterministic recursive neural networks, were developed to tackle this problem for the case of acyclic graphs (see [113] and references therein). Despite some limitations, such networks can be used for conducting QSAR studies for congeneric data sets with acyclic substituents. And, indeed, one such network, the recursive cascade correlation neural network (RCCNN), belonging to the aforementioned second type, has successfully been used by Micheli et al. for building QSAR models for benzodiazepines [49]. One can expect that, in the future, with further developments in this field, graph-based ANNs will be widely accepted in QSAR studies and maybe even revolutionize this area.

3. Conclusions

In this paper, we considered only a part of application fields of ANNs in QSAR studies. The untouched issues include the unsupervised learning (clustering, data dimensionality reduction, self-organizing maps, alignment optimization), classification (special pattern recognition neural network architectures, virtual screening with neural classificators, methods of rule extraction and knowledge integration, fuzzy neural networks), multiobjective learning, inverse task, QSCAR, and many other important topics. The question arises: What is the cause of such unprecedented creativity of researchers working in this field? Why are ANNs so popular in this research domain and their popularity grows from year to year (see Fig. 8.1).

ANNs are not the only "model-free mapping device capable of approximating any nonlinear function" [63] known to QSAR community. Some other machine learning approaches, such as the support vector machines, can do the same. In addition, the nonlinearity is not very important in many QSAR fields, and the differences in performance between ANN and non-ANN approaches are often marginal. The answer seems to be as follows. To develop a new statistical approach or, at least, deeply understand the related modern mathematical papers, it is necessary to be a good mathematician. On the other hand, many chemists and biologists can not only understand the structure and operations of neural networks but also be very creative in this field. Therefore, the naturalness, simplicity, and clarity of neural networks in comparison with many alternative machine learning approaches attracts many scientists to this area and promotes their high creativity in it. That is why one can expect that the rule of the linear growth of the number of ANN applications in QSAR and related areas, which was uncovered in the beginning of this chapter and shown in Fig. 8.1, will hold true at least in the nearest future.

References

1. Hiller SA, Glaz AB, Rastrigin LA, Rosenblit AB (1971) Recognition of physiological activity of chemical compounds on perceptron with random adaptation of structure. Dokl Akad Nauk SSSR 199:851–853.
2. Minsky M, Papert S (1969) Perceptrons. MIT Press, Cambridge, MA.
3. Hiller SA, Golender VE, Rosenblit AB, Rastrigin, LA, Glaz AB (1973) Cybernetic methods of drug design. I. Statement of the problem—the perceptron approach. Comp. Biomed. Res 6:411–421.
4. Aoyama T, Suzuki Y, Ichikawa H (1990) Neural networks applied to structure-activity relationships. J Med Chem 33:905–908.
5. Aoyama T, Suzuki,Y, Ichikawa H (1990) Neural networks applied to pharmaceutical problems. III. Neural networks applied to quantitative structure-activity relationship (QSAR) analysis. J Med Chem 33:2583–2590.
6. Zupan J, Gasteiger J (1999) Neural networks in chemistry. Wiley-VCH, Weinheim.
7. Devillers J (ed) (1996) Neural networks in QSAR and drug design. Academic Press, San Diego, CA.
8. Manallack DT, Livingstone DJ (1995) Neural networks and expert systems in molecular design. Methods and Principles in Medicinal Chemistry 3:293–318.
9. Winkler DA, Maddalena DJ (1995) QSAR and neural networks in life sciences. Series in Mathematical Biology and Medicine 5:126–163.
10. Maddalena DJ (1996) Applications of artificial neural networks to quantitative structure-activity relationships. Expert Opinion on Therapeutic Patents 6:239–251.
11. Anzali S, Gasteiger J, Holzgrabe U, Polanski J, Sadowski J, Teckentrup A, Wagener M (1998) The use of self-organizing neural networks in drug design. Persp Drug Disc Des 9-11:273–299.
12. Schneider G, Wrede P (1998) Artificial neural networks for computer-based molecular design. Progress in Biophysics and Molecular Biology 70:175–222.
13. Kovesdi I, Dominguez-Rodriguez MF, Orfi L, Naray-Szabo G, Varro A, Papp JG, Matyus P (1999) Application of neural networks in structure-activity relationships. Med Res Rev 19:249–269.
14. Manallack DT, Livingstone DJ (1999) Neural networks in drug discovery: have they lived up to their promise? Eur J Med Chem 34:195–208.

15. Ochoa C, Chana A, Stud M (2001) Applications of neural networks in the medicinal chemistry field. Curr Med Chem 1:247–256.
16. Terfloth L, Gasteiger J (2001) Neural networks and genetic algorithms in drug design. Drug Discovery Today 6:S102–S108.
17. Winkler DA, Burden FR (2002) Application of neural networks to large dataset QSAR, virtual screening, and library design. Methods in Mol Biol 201:325–367.
18. Halberstam NM, Baskin I I, Palyulin VA, Zefirov NS (2003) Neural networks as a method for elucidating structure-property relationships for organic compounds. Russ Chem Rev 72:629–649.
19. Kaiser KLE (2003) The use of neural networks in QSARs for acute aquatic toxicological endpoints. J Mol Struct (Theochem) 622:85–95.
20. Kaiser KLE (2003) Neural networks for effect prediction in environmental and health issues using large datasets. QSAR Comb Sci 22:185–190.
21. Livingstone D J, Manallack DT (2003) Neural networks in 3D QSAR. QSAR Comb Sci 22:510–518.
22. Niculescu SP (2003) Artificial neural networks and genetic algorithms in QSAR. J Mol Struct (Theochem) 622:71–83.
23. Novic M, Vracko M (2003) Artificial neural networks in molecular structures-property studies. Data Handling in Science and Technology 23:231–256.
24. Polanski J (2003) Self-organizing neural networks for pharmacophore mapping. Advanced Drug Delivery Reviews 55:1149–1162.
25. Taskinen J, Yliruusi J (2003) Prediction of physicochemical properties based on neural network modeling. Advanced Drug Delivery Reviews 55:1163–1183.
26. Winkler DA (2004) Neural networks as robust tools in drug lead discovery and development. Molecular Biotechnology 27:138–167.
27. Winkler DA, Burden FR (2004) Bayesian neural nets for modeling in drug discovery. Drug Discovery Today 2:104–111.
28. Shoji R (2005) The potential performance of artificial neural networks in QSTRs for predicting ecotoxicity of environmental pollutants. Curr. Comput-Aided Drug Des 1:65–72.
29. Livingstone DJ, Hesketh G, Clayworth D (1991) Novel method for the display of multivariate data using neural networks. J Mol Graph 9:115–118.
30. Rose VS, Croall IF, MacFie HJH (1991) An application of unsupervised neural network methodology (Kohonen topology-preserving mapping) to QSAR analysis. QSAR 10:6–15.
31. Peterson KL (1992) Counter-propagation neural networks in the modeling and prediction of Kovats indexes for substituted phenols. Anal Chem 64:379–386.
32. Liu Q, Hirono S, Moriguchi I (1992) Comparison of the functional-link net and the generalized delta rule net in quantitative structure-activity relationship studies. Chem Pharm Bull 40:2962–2969.
33. Baskin II, Palyulin VA, Zefirov NS (1993) Methodology of searching for direct correlations between structures and properties of organic compounds by using computational neural networks. Dokl Akad Nauk 333:176–179.
34. Baskin II, Palyulin VA, Zefirov NS (1997) A neural device for searching direct correlations between structures and properties of chemical compounds. J Chem Inf Comput Sci 37:715–721.
35. Lohninger H (1993) Evaluation of neural networks based on radial basis functions and their application to the prediction of boiling points from structural parameters. J Chem Inf Comput Sci 33:736–744.
36. Brinn M, Walsh PT, Payne MP, Bott B (1993) Neural network classification of mutagens using structural fragment data. SAR QSAR Env Res 1:169–210.
37. Tetko IV, Luik AI, Poda GI (1993) Applications of neural networks in structure-activity relationships of a small number of molecules. J Med Chem 36:811–814.
38. Sutter JM, Dixon SL, Jurs PC (1995) Automated descriptor selection for quantitative structure-activity relationships using generalized simulated annealing. J Chem Inf Comput Sci 35:77–84.

39. So S-S, Karplus M (1996) Evolutionary optimization in quantitative structure-activity relationship: an application of genetic neural networks. J Med Chem 39:1521–1530.
40. Viswanadhan VN, Mueller GA, Basak SC, Weinstein JN (1996) A new QSAR algorithm combining principal component analysis with a neural network: application to calcium channel antagonists. Network Science 2.
41. Domine D, Devillers J, Wienke D, Buydens L (1997) ART 2-A for optimal test series design in QSAR. J Chem Inf Comput Sci 37:10–17.
42. Sato K, Nakagawa J, Matsuzaki H (1997) Bayesian neural network approach to quantitative structure-activity relationships in carboquinones. Ann Rept Tohoku College of Pharmacy 44:187–193.
43. Kaiser KLE, Niculescu SP, McKinnon MB (1997) On simple linear regression, multiple linear regression, and elementary probabilistic neural network with Gaussian kernel's performance in modeling toxicity values to fathead minnow based on Microtox data, octanol/water partition coefficient, and various structural descriptors for a 419-compound dataset. In: Quantitative structure-activity relationships in environmental sciences, VII, Proceedings of QSAR 96, Elsinore, Denmark, June 24–28, 1996, pp. 285–297.
44. Polanski J (1997) The receptor-like neural network for modeling corticosteroid and testosterone binding globulins. J Chem Inf Comput Sci 37:553–561.
45. Kovalishyn VV, Tetko IV, Luik AI, Kholodovych VV, Villa AEP, Livingstone DJ (1998) Neural network studies 3. Variable selection in the cascade-correlation learning architecture. J Chem Inf Comput Sci 38:651–659.
46. Zupan J, Novic M (1999) Optimization of structure representation for QSAR studies. Anal Chim Acta 388:243–250.
47. Espinosa G, Yaffe D, Cohen Y, Arenas A, Giralt F (2000) Neural network based quantitative structural property relations (QSPRs) for predicting boiling points of aliphatic hydrocarbons. J Chem Inf Comput Sci 40:859–879.
48. Li, P, Cheng Y-Y (2000) Studies on quantitative structure-activity relationships of benzodiazepines using fuzzy neural networks. Gaodeng Xuexiao Huaxue Xuebao 21:1473–1478.
49. Micheli A, Sperduti A, Starita A, Bianucci AM (2001) Analysis of the internal representations developed by neural networks for structures applied to quantitative structure-activity relationship studies of benzodiazepines. J Chem Inf Comput Sci 41:202–218.
50. Tetko IV, Kovalishyn VV, Livingstone DJ (2001) Volume learning algorithm artificial neural networks for 3D QSAR studies. J Med Chem 44:2411–2420.
51. Polanski J, Gieleciak R, Bak A (2002) The comparative molecular surface analysis (CoMSA)—a nongrid 3D QSAR method by a coupled neural network and PLS system: predicting pK_a values of benzoic and alkanoic acids. J Chem Inf Comput Sci 42:184–191.
52. Patankar SJ, Jurs PC (2002) Prediction of glycine/NMDA receptor antagonist inhibition from molecular structure. J Chem Inf Comput Sci 42:1053–1068.
53. Mosier PD, Jurs PC (2002) QSAR/QSPR studies using probabilistic neural networks and generalized regression neural networks. J Chem Inf Comput Sci 42:1460–1470.
54. Espinosa G, Arenas A, Giralt F (2002) An integrated SOM-fuzzy ARTMAP neural system for the evaluation of toxicity. J Chem Inf Comput Sci 42:343–359.
55. Agrafiotis DK, Cedeno W (2002) Feature selection for structure-activity correlation using binary particle swarms. J Med Chem 45:1098–1107.
56. Izrailev S, Agrafiotis DK (2002) Variable selection for QSAR by artificial ant colony systems. SAR QSAR Env Res 13:417–423.
57. Arakawa M, Hasegawa K, Funatsu K (2003) Application of the novel molecular alignment method using the Hopfield neural network to 3D-QSAR. J Chem Inf Comput Sci 43:1396–1402.
58. Hemmateenejad B, Akhond M, Miri R, Shamsipur M (2003) Genetic algorithm applied to the selection of factors in principal component-artificial neural networks: application to QSAR study of calcium channel antagonist activity of 1,4-dihydropyridines (nifedipine analogs). J Chem Inf Comput Sci 43:1328–1334.

59. Baskin II, Halberstam NM, Artemenko NV, Palyulin VA, Zefirov NS (2003) NASAWIN—a universal software for QSPR/QSAR studies. In: Ford M et al. (eds) EuroQSAR 2002 designing drugs and crop protectants: processes, problems and solutions. Blackwell Publishing, pp. 260–263.
60. Cedeno W, Agrafiotis DK (2003) Application of niching particle swarms to QSAR and QSPR. In: Ford M et al. (eds) EuroQSAR 2002 designing drugs and crop protectants: processes, problems and solutions . Blackwell Publishing, pp. 255–259.
61. Bayram E, Santago P II, Harris R, Xiao Y-D, Clauset AJ, Schmitt JD (2004) Genetic algorithms and self-organizing maps: a powerful combination for modeling complex QSAR and QSPR problems. J Comp-Aided Mol Des 18:483–493.
62. Baurin N, Mozziconacci J-C, Arnoult E, Chavatte P, Marot C, Morin-Allory L (2004) 2D QSAR consensus prediction for high-throughput virtual screening. An application to COX-2 inhibition modeling and screening of the NCI. J Chem Inf Comput Sci 44:276–285.
63. Maggiora GM, Elrod DW, Trenary RG (1992) Computational neural networks as model-free mapping devices. J Chem Inf Comput Sci 32:732–741.
64. Kolmogorov AN (1957) On the representation of continuous functions of many variables by superposition of continuous functions of one variable and addition. Dokl Akad Nauk SSSR 114:953–956.
65. Kurková V (1992) Kolmogorov's theorem and multilayer neural networks. Neural Networks 5:501–506.
66. Manallack DT, Livingstone DJ (1994) Limitations of functional-link nets as applied to QSAR data analysis. Quantitative Structure-Activity Relationships 13:18–21.
67. Stuetzle W, Mittal Y (1979) Some comments on the asymptotic behavior of robust smoothers. In: Gasser T, Rosenblatt M (eds) Smoothing techniques for curve estimation. Springer-Verlag, Helderberg.
68. Tetko IV (2002) Neural network studies, 4. Introduction to associative neural networks. J Chem Inf Comput Sci 42:717–728.
69. Raevsky OA, Trepalin SV, Trepalina HP, Gerasimenko VA, Raevskaja OE (2002) SLIPPER-2001—software for predicting molecular properties on the basis of physicochemical descriptors and structural similarity. J Chem Inf Comput Sci 42:540–549.
70. Féraud R, Clérot F (2002) A methodology to explain neural network classification. Neural Networks 15:237–246.
71. Guha R, Stanton DT, Jurs PC (2005) Interpreting computational neural network quantitative structure-activity relationship models: a detailed interpretation of the weights and biases. J Chem Inf Mod 45:1109–1121.
72. Aoyama, T, Ichikawa H (1992) Neural networks as nonlinear structure-activity relationship analyzers. Useful functions of the partial derivative method in multilayer neural networks. J Chem Inf Comput Sci 32:492–500.
73. Baskin II, Ait AO, Halberstam NM, Palyulin VA, Zefirov NS (2002) An approach to the interpretation of backpropagation neural network models in QSAR studies. SAR QSAR Env Res 13:35–41.
74. Halberstam NM, Baskin II, Palyulin VA, Zefirov NS (2002) Quantitative structure-conditions-property relationship studies. Neural network modeling of the acid hydrolysis of esters. Mendeleev Commun 12:185–186.
75. Guha R, Jurs PC (2005) Interpreting computational neural network QSAR models: a measure of descriptor importance. J Chem Inf Mod 45:800–806.
76. Stanton DT (2003) On the physical interpretation of QSAR models. J Chem Inf Comput Sci 43:1423–1433.
77. Manallack D, Livingstone DJ (1992) Artificial neural networks: application and chance effects for QSAR data analysis. Med Chem Res 2:181–190.
78. Livingstone DJ, Salt DW (1992) Regression analysis for QSAR using neural networks. Bioorg Med Chem Let 2:213–218.
79. Livingstone DJ, Manallack DT (1993) Statistics using neural networks: chance effects. J Med Chem 36:1295–1297.

80. Tetko IV, Livingstone DJ, Luik AI (1995) Neural network studies, 1. Comparison of overfitting and overtraining. J Chem Inf Comput Sci 35:826–833.
81. Andrea TA, Kalayeh H (1991) Applications of neural networks in quantitative structure-activity relationships of dihydrofolate reductase inhibitors. J Med Chem 34:2824–2836.
82. Ezhov AA, Shumsky SA (1998) Neurocomputing and its applications in economics and business. MIPhI, Moscow.
83. Rissanen J (1983) A universal prior for the integers and estimation by minimum description length. Annals of Statistics 11:416–431.
84. Tikhonov AN, Arsenin VY (1977) Solutions of ill-posed problems. Winston & Sons, Washington, DC.
85. Burden FR, Winkler DA (1999) Robust QSAR models using Bayesian regularized neural networks. J Med Chem 42:3183–3187.
86. Agrafiotis DK, Cedeño W, Lobanov VS (2002) On the use of neural network ensembles in QSAR and QSPR. J Chem Inf Comput Sci 42:903–911.
87. Whitley DC, Ford MG, Livingstone DJ (2000) Unsupervised forward selection: a method for eliminating redundant variables. J Chem Inf Comput Sci 40:1160–1168.
88. Tetko IV, Villa AEP, Livingstone DJ (1996) Neural network studies, 2. Variable Selection. J Chem Inf Comput Sci 36:794–803.
89. Artemenko NV, Baskin II, Palyulin VA, Zefirov NS (2003) Artificial neural network and fragmental approach in prediction of physicochemical properties of organic compounds. Russ Chem Bull 52:20–29.
90. Gustafsson MG. (2005) Independent component analysis yields chemically interpretable latent variables in multivariate regression. J Chem Inf Comput Sci 45:1244–1255.
91. Hansch C, Fujita T (1964) σ-ρ-π Analysis. A method for the correlation of biological activity and chemical structure. J Am Chem Soc 86:1616–1626.
92. Free SM Jr, Wilson JW (1964) A mathematical contribution to structure-activity studies. J Med Chem 7:395–399.
93. Kvasnička V, Sklenák Š, Pospichal J (1993) Neural network classification of inductive and resonance effects of substituents. J Am Chem Soc 115:1495–1500.
94. Baskin II, Keschtova SV, Palyulin VA, Zefirov NS (2000) Combining molecular modeling with the use of artificial neural networks as an approach to predicting substituent constants and bioactivity. In: Gundertofte K, Jørgensen FS (eds) Molecular modeling and prediction of bioactivity. Kluwer Academic/Plenum Publishers, New York, pp. 468–469.
95. Chiu T-L, So S-S (2004) Development of neural network QSPR models for Hansch substituent constants, 1. Method and validations. J Chem Inf Comput Sci 44:147–153.
96. Chiu T-L, So S-S (2004) Development of neural network QSPR models for Hansch substituent constants, 2. Applications in QSAR studies of HIV-1 reverse transcriptase and dihydrofolate reductase inhibitors. J Chem Inf Comput Sci 44:154–160.
97. Baskin II, Halberstam NM, Mukhina TV, Palyulin VA, Zefirov NS (2001) The learned symmetry concept in revealing quantitative structure-activity relationships with artificial neural networks. SAR QSAR Env Res 12:401–416.
98. Baskin II, Skvortsova MI, Stankevich IV, Zefirov NS (1994) Basis of invariants of labeled molecular graphs. Doklady Chemistry 339:231–234.
99. Baskin II, Skvortsova MI, Stankevich IV, Zefirov NS (1995) On the basis of invariants of labeled molecular graphs. J Chem Inf Comput Sci 35:527–531.
100. Baskin II, Palyulin VA, Zefirov NS (1993) Computational neural networks as an alternative to linear regression analysis in studies of quantitative structure-property relationships for the case of the physicochemical properties of hydrocarbons. Dokl Akad Nauk 332:713–716.
101. Artemenko NV, Baskin II, Palyulin VA, Zefirov NS (2001) Prediction of physical properties of organic compounds by artificial neural networks in the framework of substructural approach. Doklady Chemistry 381:317–320.
102. Artemenko NV, Palyulin VA, Zefirov NS (2002) Neural-network model of the lipophilicity of organic compounds based on fragment descriptors. Doklady Chemistry 383:114–116.

103. Zhokhova NI, Baskin II, Palyulin VA, Zefirov AN, Zefirov NS (2003) Fragmental descriptors in QSPR: flash point calculations. Russ Chem Bull 52:1885–1892.
104. Zhokhova NI, Baskin II, Palyulin VA, Zefirov AN, Zefirov NS (2003) Calculation of the enthalpy of sublimation by the QSPR method with the use of a fragment approach. Russ J Appl Chem 76:1914–1919.
105. Zhokhova NI, Baskin II, Palyulin VA, Zefirov AN, Zefirov NS (2004) Fragment descriptors in QSPR: application to magnetic susceptibility calculations. J Structural Chem 45:626–635.
106. Liubimova IK, Abilev SK, Gal'berstam NM, Baskin II, Paliulin VA, Zefirov NS (2001) Computer-aided prediction of the mutagenic activity of substituted polycyclic compounds. Biology Bull 28:139–145.
107. Radchenko EV, Baranova OD, Palyulin VA, Zefirov NS (2003) Non-linear molecular field topology analysis by means of artificial neural networks. In: Ford M et al. (eds) EuroQSAR 2002 designing drugs and crop protectants: processes, problems and solutions. Blackwell Publishing, pp. 317–318.
108. Cramer RD, Patterson DE, Bunce JD (1988) Comparative molecular field analysis (CoMFA), 1. Effect of shape and binding of steroids to carrier proteins. J Am Chem Soc 110:5959–5967.
109. Hasegawa K, Matsuoka S, Arakawa M, Funatsu K (2002) New molecular surface-based 3D-QSAR method using Kohonen neural network and three-way PLS. Comput Chem 26:583–589.
110. Baskin II, Tikhonova IG, Palyulin VA, Zefirov NS (2006) A combined approach to building 3D-QSAR models. In QSAR and Molecular Modelling in Rational Design of Bioactive Molecules, Proceedings of the European Symposium on Structure-Activity Relationships (QSAR) and Moleular Modelling, 15th, Istanbul, Turkey, Sept. 5-10, 2004/Eds by E. Aki (Şener) and I. Yalçin, pp. 123–124.
111. Kireev DB (1995) ChemNet: a novel neural network based method for graph/property mapping. J Chem Inf Comput Sci 35:175–180.
112. Ivanciuc O (2001) Molecular structure encoding into artificial neural networks topology. Roumanian Chem Quart Rev 8:197–220.
113. Baldi P, Rosen-Zvi M (2005) On the relationship between deterministic and probabilistic directed graphical models: from Bayesian networks to recursive neural networks. Neural Networks 18:1080–1086.