Russian Chemical Reviews **72** (7) 629 – 649 (2003)

DOI 10.1070/RC2003v072n07ABEH000754

Neural networks as a method for elucidating structure – property relationships for organic compounds

N M Halberstam, I I Baskin, V A Palyulin, N S Zefirov

Contents

Abstract. The published data devoted to the use of the neural network approach in the simulation of structure – property relationships for organic compounds are reviewed. The basic principles of the neural network simulation are discussed along with the characteristic features of the neural network approach typical of the representation and classification of structural chemical data. Brief information on neural network models of spectral characteristics, reactivities, physicochemical properties and biological activities of organic compounds is presented. The bibliography includes 159 references.

I. Introduction

The use of computer technologies in virtually all branches of science has really acquired a mass character in the past decade. The wealth of experimental data accumulated by now makes it possible to focus specific attention on the methods for generalisation and mathematical processing of diverse parameters of known compounds aimed at the computer simulation and forecasting of properties of novel compounds that have not been synthesised nor investigated hitherto. This, in turn, opens up broad opportunities for the solution of one of the central tasks of chemical science, *viz*., a purposeful search for novel compounds and materials with predetermined properties.

Artificial neural networks have become one of the most popular methods for the construction of various quantitative relationships. The methodology of artificial neural networks, which enables the construction of nonlinear models of any degree of complexity, has found wide application in a search for quantitative relationships between structures of organic compounds and their physicochemical properties (QSPR) or biological activities (QSAR).

N M Halberstam, I I Baskin, V A Palyulin, N S Zefirov N D Zelinsky Institute of Organic Chemistry, Russian Academy of Sciences, Leninsky prosp. 47, 119991 Moscow, Russian Federation. Fax (7-095) 135 53 28. Tel. (7-095) 939 35 57. E-mail: n.halb.@g23.relcom.ru (N M Halberstam)
Department of Chemistry, M V Lomonosov Moscow State University, Leninskie Gory, 119992 Moscow, Russian Federation.
Fax (7-095) 939 31 81. Tel. (7-095) 939 35 57.
E-mail: baskin@org.chem.msu.su (I I Baskin), vap@org.chem.msu.su (V A Palyulin). Tel. (7-095) 939 16 20. E-mail: zefirov@org.chem.msu.su (N S Zefirov)

Received 23 July 2002
Uspekhi Khimii **72** (7) 706 – 727 (2003); translated by R L Birnova

Among the known architectures of neural networks used for elucidating structure – property relationships (hereinafter, the term 'property' relates to both physicochemical properties and biological activities of organic compounds), multilayered feedforward backpropagation neural networks come first in popularity. Their attractiveness is mostly due to the ability of such neural networks to generalise and approximate data with high accuracy and to the possibility of processing large arrays of disembodied data.

A search for neural network structure – property relationships may be reduced to sequential execution of the following operations: (1) description of structures of compounds under study by special numerical parameters (descriptors); (2) choice of an optimum set of descriptors; (3) classification of the whole set of experimental data into training and control samples; (4) selection of the most adequate type, architecture and the method for training of the neural network; (5) choice of statistic parameters for estimating the quality of training; (6) training of neural networks and estimation of training results; (7) the use of neural network models for forecasting the properties and activities of hitherto unknown compounds.

The first publications devoted to the application of neural networks (perceptrons) to the solution of various chemical problems appeared in the late 1980's.[1] The interest of chemists in the new method has begun to increase sweepingly since that time.

Neural networks, commonly referred to as artificial neural networks or computing neural networks, represent a simplified mathematical model for information processing by the human brain. However, the majority of existing architectures of neural networks cannot reproduce brain functions exactly and should be considered as a modification of parallel algorithms.[2, 3]

Owing to their ability for being trained, combining and generalising information, neural networks have been successfully used in chemistry, especially in those cases where the analytical types of relationships between the structures and properties of chemical compounds are unknown.[4]

II. Basic principles of neural network simulation

1. Description of architectures and methods for training of neural networks used for elucidating structure – property relationships

Neural networks consist, as a rule, of a set of relatively simple computing elements termed neurons; the latter are arranged in several layers and are linked to one another by numerous connections. The layers are usually subdivided into three groups, *viz*., input, hidden and output (Fig. 1). Each connection between two neurons is defined by a real number termed connection

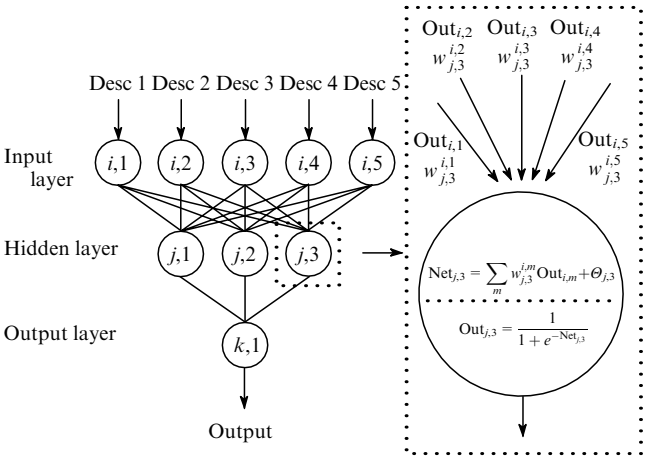


Figure 1. The structure of a feedforward neural network. The enlarged fragment depicts the main computational operations performed by the neuron with a sigmoidal transfer function. Desc is descriptor.

weight. The sum of connection weighting coefficients of interneural bonds determines the computational capabilities of neural networks; therefore, training of neural networks is reduced to the adjustment of their connection weighting coefficients.

The computations within each *i*th neuron of the *n*th layer are performed in two steps (see Fig. 1), *viz*., computation of the net input signal Net_j by summation of all weighted input signals and computation of the output signal Out_j of this neuron on the basis of the transfer function:

$$Net_j = \sum_j w_{ij} Out_i + \Theta_j,$$

$$Out_j = f(Net_j),$$

where Net_j is the net input signal of the neuron j pertaining to the layer n; Out_i is the output signal of the neuron i pertaining to the layer $n-1$; w_{ij} is the weighting coefficient of the connection between the neurons i and j; Θ_j is the threshold value for the neuron j; and Out_j is the output signal of the neuron j in the layer n.

The sigmoidal functions (1) are used most commonly as transfer functions, but threshold (2), linear (3), hyperbolic tangential (4) and some other functions can also be used.[5,6]

$$Out_j = \frac{1}{1 + e^{-Net_j}}, \quad (1)$$

$$Out_j = \begin{cases} 0, & Net_j < 0 \\ 1, & Net_j \geqslant 0 \end{cases}, \quad (2)$$

$$Out_j = k\, Net_j \quad (k = \text{const}), \quad (3)$$

$$Out_j = \frac{e^{Net_j} - e^{-Net_j}}{e^{Net_j} + e^{-Net_j}}. \quad (4)$$

The so-called bias pseudoneurons[2] with a constant output signal equal to 1 are usually added to each (except for the output layer) layer of a neural network instead of the threshold values Θ.

Prior to training, all connection weights are initialised by random numbers. A correct choice of initialisation boundaries enables reduction of the training time and improvement of the quality of the neural network models derived.[7,8] In some cases, where the training procedures (*i.e*., the adjustment of connection weights) do not converge, the training is repeated using another set of parameters or a different mode of initialisation of connection weights.

The architecture of a neural network is determined by the connection topology between the neurons. Simulation of structure–property relationships usually employs multilayered neural networks the number of layers in which is determined by a specific architecture of the neural network.[2,9] Each neuron pertaining to one layer has the same number of input connections with the preceding layer. The output signals of the last computing layer of the network represent computed output values of the whole network. When neural networks are used for forecasting the properties of chemical compounds, the values of output signals of the input neurons are set up with regard for the values of computed descriptors after normalisation (see below); the signals identified with the values of predicted properties (also, taking normalisation into account) are read from the output neurons.

a. Multilayered feedforward neural networks

The majority of publications devoted to the elucidation of structure–property relationships deal with multilayered feedforward neural networks. The main advantages of such networks include their ability to establish multiparametric nonlinear relationships with a high interpolation accuracy even in those cases where experimental data are not representative enough or contain noises.[10]

A characteristic feature of feedforward neural networks is a layer-by-layer transfer of signals from the input of the network to its output. Classical feedforward neural networks usually possess several hidden layers (Fig. 2 *a*) or contain exclusively an input and an output layer (Fig. 2 *b*).

The adjustment of weighting coefficients of the connections during training of multilayered feedforward neural networks is carried out sequentially starting from the output layer connections; therefore, the methodology of training of such networks has got the name of back propagation of errors.[11,12] A brief description of the most popular procedures is given below.

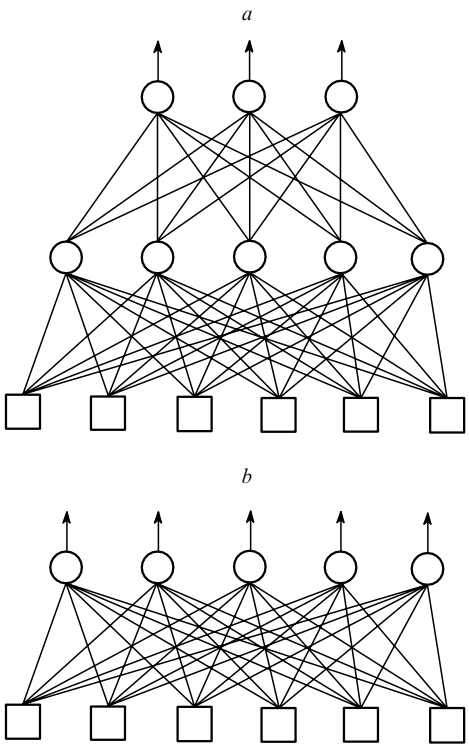


Figure 2. The classical feedforward neural networks with one hidden layer (*a*) and without any hidden layers (*b*).

b. A generalised delta rule

Training of each layer n out of N layers of a neural network is carried out in accordance with the Widrow – Hoff rule or the delta rule:[13, 14]

$$\Delta w_{ij} = \eta\, \delta_j\, o_i\,,$$

where Δw_{ij} is the change in the weighting coefficient of the connection between the neurons i and j; η is the empirical constant for the training rate; δ_j is the computational error for the neuron j pertaining to the layer n and o_i is the output signal of the neuron i pertaining to the layer $n-1$.

In the case of a generalised delta rule, an additional parameter termed a 'training momentum' and calculated with consideration of changes in the weighting coefficients in the previous iteration is introduced in order to avoid oscillations which often take place where error surface is characterised by a very narrow valley region:[13 – 15]

$$\Delta w_{ij}^{(k+1)} = \eta\, \delta_j\, o_i + \alpha\, \Delta w_{ij}^{(k)},$$

$$\delta_j = \begin{cases} f'(\mathrm{Net}_j)(t_j - o_j), & n = N \\ f'(\mathrm{Net}_j) \sum\limits_m \delta_m w_{jm}, & n \neq N, \end{cases}$$

where t_j and o_j are the experimental and the calculated values of the property under investigation for the neuron j, respectively; α is the constant reflecting the influence of the training momentum and k is the iteration number.

c. The resilient propagation method (RPROP)

If the training is performed with the use of RPROP, the adjustment of connection weighting coefficients is performed exclusively on the basis of the data about the signs of partial derivatives of error functions of the neural network E usually defined as a sum of squared errors on the output neurons.[16, 17] The changes in the connection weighting coefficients are calculated in the following way:

$$\Delta w_{ij}^{(t)} = \begin{cases} -\Delta_{ij}^{(t)}, \text{ if } \dfrac{\partial E^{(t)}}{\partial w_{ij}} > 0 \\ +\Delta_{ij}^{(t)}, \text{ if } \dfrac{\partial E^{(t)}}{\partial w_{ij}} < 0 \\ \dfrac{\partial E^{(t)}}{\partial w_{ij}} = 0\,, \end{cases}$$

$$\Delta_{ij}^{(t)} = \begin{cases} \eta^+ \cdot \Delta_{ij}^{(t-1)}, \text{ if } \dfrac{\partial E^{(t-1)}}{\partial w_{ij}} \cdot \dfrac{\partial E^{(t)}}{\partial w_{ij}} > 0 \\ \eta^- \cdot \Delta_{ij}^{(t-1)}, \text{ if } \dfrac{\partial E^{(t-1)}}{\partial w_{ij}} \cdot \dfrac{\partial E^{(t)}}{\partial w_{ij}} < 0 \\ \dfrac{\partial E^{(t-1)}}{\partial w_{ij}} \cdot \dfrac{\partial E^{(t)}}{\partial w_{ij}} = 0. \end{cases}$$

Here, $\dfrac{\partial E^{(t)}}{\partial w_{ij}}$ is the value of the partial derivative of the error function of the neural network with respect to the connection weight w_{ij} in the ith iteration, η^+ and η^- are the empirical constants for the increase or decrease in the training rate and t is the iteration number.

d. Quasi-Newtonian training methods

This group of methods is based on the Newtonian principle of function fitting. A vector all the components of which are taken equal to the partial derivatives of the error function of the network with respect to all weighting coefficients is calculated for each iteration. The Broyden – Fletcher – Goldfarb – Shanno (BFGS) method is the most efficient in this group.[14, 18]

2. Other neural network architectures

Neural networks having different architectures are successfully used for the simulation of structure – property relationships. A brief description of some of these architectures is given below.

Feedforward neural networks containing at least one hidden layer may possess additional direct connections linking the neurons of the input and output layers [19 – 21] as well as recursive bonds between the output and the input of a neuron. [21, 22]

Kvasnicka *et al.*[23, 24] have proposed the use of a tree-shaped neural network for the solution of problems related to classification of compounds; the topology of such a network was the most close to the molecular structures under investigation. The values 0 or 1 were ascribed to the input neurons so as to obtain subgraphs isomorphous to the structures of the molecules. Later, a similar approach was realised in the computer programme 'ChemNet'.[25]

An interesting neural network procedure for the description of chemical compounds has been proposed, which allows elucidation of structure – property relationships without computation of descriptors for molecules as a whole using only atom and bond descriptors.[26, 27] In this method, the presentation of structures is effected through the use of an additional neural network which consists of two principal elements, *viz*., a sensorial area to which the basic structural information is fed and an 'eye' which transforms the data from the sensorial area into a signal invariant to the renumbering of the atoms in compounds under study. The structural information thus processed is fed to the input of ordinary feedforward neural networks in order to obtain structure – property relationships.

The Kohonen neural networks (see Refs 28 – 32) (Fig. 3 *a*) widely employed in cluster analysis allow data mapping in such a way that similar vectors of input values are mapped onto neighbouring output neurons on the lattice.

Counterpropagation neural networks [33, 34] utilise two different types of layers, *viz*., a hidden Kohonen layer and an output Grossberg layer (Fig. 3 *b*). The salient advantages of networks trained by counterpropagation include a relatively small (of the order of several hundreds) number of interactions required for the training of networks [35, 36] and the possibility of finding a global minimum of the error function for any starting setup of the weighting coefficients.[37, 38] One disadvantage is a weaker approximating capacity of such networks in comparison with other architectures.

A neural network with a linear transfer function for the output-layer neurons and hidden neurons which accomplish radial-basic nonlinear approximation is usually referred to as a radial-basic neural network. This is also used for elucidating structure – property relationships.[39 – 41]

The functioning of Fuzzy ARTMAP (Fuzzy Adaptive Resonance Theory for Mapping) neural networks is based on clusterisation (categorisation) of training set vectors in accordance with the adaptive resonance theory (see Ref. 42 and References cited therein). In this method, categorisation of vectors is performed by comparing each successive vector to reference vectors describing previously established categories (clusters). If a successive vector 'resembles' a reference vector in terms of a definite proximity criterion, it is used for its adjustment. In the opposite case, a vector becomes a representative of a new data category and is memorised as a new reference vector. This procedure is implemented in a neural network consisting of a reference layer, which estimates the 'similarity' of the vectors, a recognising layer where each neuron describes its own data category (cluster) and several accessory elements. Such architectures have been termed ART-1 (categorisation of binary vectors) and ART-2 (categorisation of real number vectors). ARTMAP represents a modular neural network made up of two networks for categorisation of vectors of the ART type (in QSAR/QSPR studies, these vectors are linked to the vectors corresponding to descriptors and properties of organic compounds) and a comparison module responsible for memorisation of 'associations' between different categories of descriptors and properties.

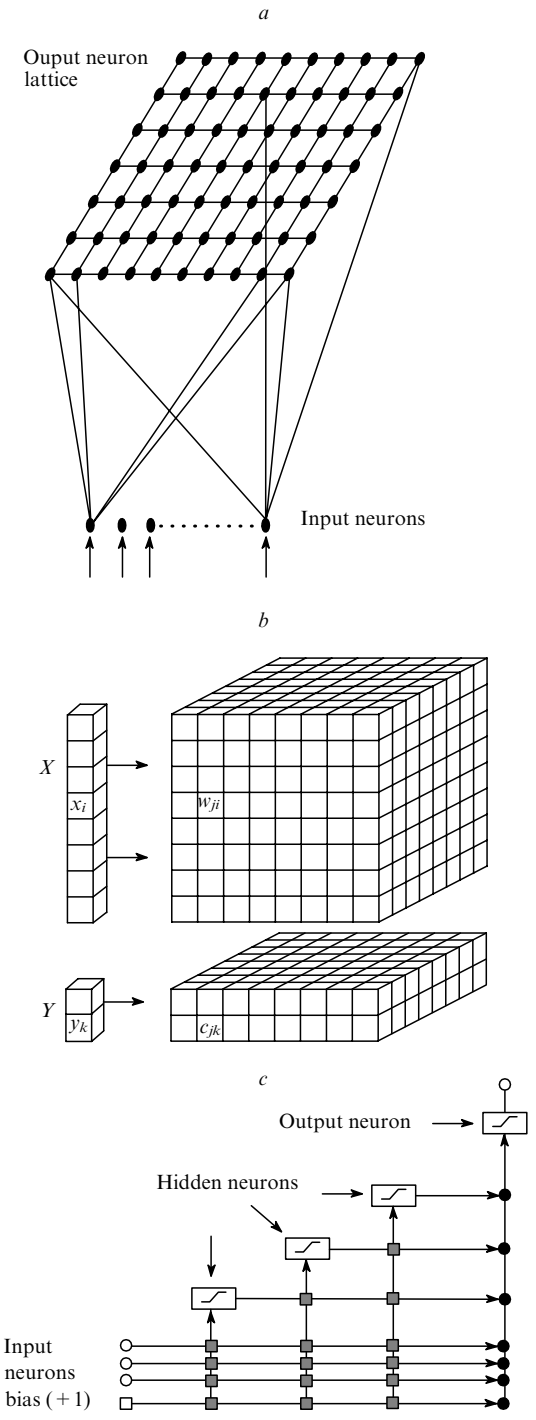


Figure 3. A schematic representation of architectures of some neural networks used for the simulation of structure–property relationships. (*a*) The Kohonen neural network, (*b*) back propagation neural network, (*c*) cascade-type neural network.

A neural network of functional links [43] is made up of two layers. The outputs of the neurons forming the input layer result from nonlinear transformations (*e.g.*, squaring, taking a square root, a sinus, *etc.*) of the values of a descriptor; each neuron in the output layer corresponds to a predicted property. It is of note that this category represents a neural network simulation of a linear regression analysis based on the use of nonlinearly transformed descriptors and properties.

Sometimes, in the simulation of structure–property relationships several neural networks having different architectures are combined into a hierarchical system.[21,44]

Alongside neural networks with static architectures (which have to be selected prior to training) there are the so-called cascade networks (Fig. 3 *c*); their topology is formed during training.[45–47] The training is begun with a neural network containing exclusively input and output neurons. If the error of the neural network exceeds the threshold value, one hidden neuron which accumulates signals from all input and previously added hidden neurons is added after each training cycle. It may therefore be assumed that each hidden neuron forms a separate layer. After addition of a successive hidden neuron, the weights of its connections are determined only once to enable maximisation of covariation of values of output signals with the error and do not change any further in the course of training. As far as the output neurons are concerned, the weights of their connections have to be adjusted every time in order to diminish the error function of the network.

III. Approaches to structural description of compounds

The first question arising in connection with the application of neural networks, especially in the solution of various chemical problems, is related to the choice of a method for the description of source data.

The combination of parameters fed to the input of a neural network is usually presented in the form of a vector. Therefore, a topological or a three-dimensional molecular structure has to be described as a unidimensional array without considerable losses of significant information. The following conditions should be met:[36]

(1) the quantity and the type of descriptors are the same for all the structures to be described;

(2) there is a correspondence between each descriptor and a particular component in a parameter vector;

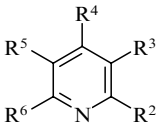
(3) different parameter vectors correspond to different molecules.

As a rule, neural network simulation of structure–property relationships utilises a set of numeric values (descriptors) [48] for the description of structures of chemical compounds containing all relevant information. According to the type of information, all descriptors are subdivided into fragmental,[49] topological,[50,51] quantum-chemical [52] and physicochemical. All of them demand additional computation prior to construction of neural network models. Moreover, one should have a clear idea about the nature of a relationship in order to select an appropriate group and a type of descriptors for each concrete case.

In the majority of studies devoted to neural network simulation of structure–property relationships, descriptors are computed with the aid of various software packages, *e.g.*, ADAPT.[53–55] This usually results in an excessive set of descriptors; the most significant ones are selected for further analysis using various procedures. Sometimes, it is rather difficult to explain the physical sense of the descriptors selected and to establish their relationship to the property under investigation.

The first attempts to use a well-known procedure for the presentation of chemical structures as modified connectivity matrices with charges on nuclei of the corresponding atoms on their diagonals and formal bond orders as non-diagonal elements were undertaken by Elrod *et al.* [56] It was proposed to form input vectors for neural networks by combining the diagonals of the connectivity matrices with subdiagonal elements taken row-by-row. These authors recommended using two connectivity matrices (one for reactants and one for reaction products) for the description of chemical reactions.[57] Evidently, the use of this approach is confined to compounds containing a small number of atoms due to a great number of elements in the connectivity matrix. In addition, such a presentation of structures is not invariant with respect to the renumbering of atoms.

The classical approach to the construction of quantitative structure – property relationships is used for the description of a series of structural analogues using substituent constants, *e.g.*, Hammett σ constants, Taft's E_s constants, lipophilic constants π, parameters L, B_1, B_2 (STERIMOL), *etc*. However, if structures contain topologically equivalent substitution positions, the use of these parameters is not quite correct. For example, the pyridine molecule contains two pairs of topologically equivalent positions of substituents, *viz*., the α-position for R^2 and R^6 and the β-position for R^3 and R^5. As a consequence, the same compound will be represented by different sets of descriptors based on the use of different parameters of substituents, which may result in ambiguous prognosis.



This problem may be overcome through the introduction into the source set of copies of compounds with different arrangements of substituents in topologically equivalent positions,[58] possessing the same activities as the starting compound. Such an enlargement of the source sample makes prognosis independent in cases of ambiguous assignment of substituents to topologically equivalent positions.

An interesting method has been proposed which enables one to reduce the number of input parameters of neural networks without any loss of significant information.[59] This approach is based on the computation of the principal components [60] which represent linear combinations of original descriptors. The values thus obtained are further used as the input parameters of neural networks.

1. Normalisation of input parameters of neural networks

Almost all publications devoted to the use of neural networks for the solution of various chemical problems describe normalisation of input data. As a rule, input parameters are normalised in such a way that their values lie in the range between 0 and 1 or are characterised by unit variance.

2. Selection of the most significant descriptors

Of no less importance is the selection, among all computed parameters, of only those parameters which make a significant contribution to the property under study. Such a selection can be made in advance or in the course of training of neural networks.

Preliminary selection of parameters is performed using independent methods, *e.g.*, stepwise multiple regression analysis,[61, 62] or neural network algorithms.[63] In the latter case, the type of neural network used for the selection of descriptors does not usually coincide with the type of main neural network used for the construction of the model.

A neural network utilising a 'genetic' training algorithm can serve as an example.[64 – 66] The 'genetic' algorithm represents a stochastic optimisation procedure based on evolutionary principles. Its operation ensures the selection of a description with the most significant parameters from the whole set of object descriptions over the course of several generations. The significance of parameters, *i.e*., the success of genetic combinations, is defined by a criterion function computed by a neural network as a residual error for the property under investigation. Diverse methods for the production of a new generation within the framework of a 'genetic' algorithm have been described. Approximation of the genetic function [67] and evolutionary programming [68] are the most popular.

Brinn *et al*.[69] have recommended using the so-called evolutionary approximation for the rejection of insignificant descriptors. Some descriptors randomly selected from a multitude of computed descriptors are fed to the input of a neural network after which the root-mean-square error is computed at the output of the network. Then, one of the input neurons is excluded from the neural network and the new root-mean-square error is recalculated. This procedure is repeated for all neurons. The descriptors are scored according to the magnitude of the change in the error at the output of the neural network. The neurons are then removed from the neural network starting from the least significant descriptors until the change in the error at the output of the network becomes significant. After replacement of discarded descriptors by new ones taken from the set of computed descriptors, the calculations are repeated.

In the course of training of neural networks, less significant descriptors can be discarded using the so-called pruning algorithms. Their main function consists of simplification of neural network architectures by discarding the connections with less significant weighting coefficients. If all the connections of input neurons are insignificant, these neurons are discarded.[70, 71] Pruning procedures can be classified into two groups,[72] *viz*., sensitivity methods based on computation of significance or sensitivity of all connections followed by elimination of the connections with the least sensitivity, and 'penalty component methods' based on the introduction of the so-called penalty functions which reduce weighting coefficients to zero values in the course of training.

Tetko *et al*.[73] have provided a detailed description of five methods designed for the computation of connection sensitivities. A method for the selection of the most significant descriptors based on the calculation of sensitivities of connections and the contribution of each descriptor to the total error of the neural network has been developed.[74, 75]

A reconstructive method based on the forgetting principle, *i.e*., a sequential decrease in the absolute values of all weighting coefficients, has been developed.[76] Reiteration of training – forgetting cycles enables enhancement of significant or weakening of less significant connections.

The degeneration algorithm of weighting coefficients is yet another example of a procedure based on penalty functions.[13, 20] In each training step, the values of calculated weighting coefficients of the connections diminish by a fraction of their previous values as a result of which insignificant weighting coefficients decrease to 0 in the course of training.

3. Choice of an optimum training set

The design of training samples is yet another problem related to the application of neural networks in chemistry. Chemical compounds constituting a training set should be evenly distributed along the whole range of possible values of input parameters.

The majority of published data deal with random division of the whole array of compounds under investigation into training and control sets, but only some of them provide a description of special procedures for the selection of an optimum training set. The use of Kohonen neural networks has been described in Refs 35, 77, 78; a neural network based on the theory of adaptive resonance (ART 2-A, see above) and intended for preliminary classification of the whole set of experimental data has been described.[79] The choice of a correct size of training set is yet another critical factor.

4. Optimisation of the number of hidden neurons

The number of adjustable parameters (weighting coefficients) of neural networks is determined by the number of hidden neurons. If the number of hidden neurons is too large and the number of adjustable parameters is comparable with the number of structures in the training set, the neural networks will try to 'memorise' the training set (the overdetermination effect). Neural networks with an insignificant number of hidden neurons are unable to utilise the whole array of input information for generalisation and classification.[80, 81]

The expediency of the introduction of a special parameter r for the estimation of an optimum number of hidden neurons has been demonstrated:[82 – 84]

$$r = \frac{N}{P},$$

where N is the number of molecules in the training set and P is the total number of connections in the neural network.

The optimum values of this parameter lie in the range $1.8 < r < 2.2$. The number of hidden neurons thus calculated enables the construction of neural networks manifesting the best forecasting ability.[80]

5. Effect of 'overtraining' of neural networks

The overdetermination of a model (see above) leads to the so-called 'overtraining' of neural networks.[85–87]

The training of neural networks includes two phases, *viz*., generalisation and memorisation of information. The first phase is characterised by a decrease in the mean computational error of the property under study for compounds in both training and control sets. However, in the next phase, memorisation, the error calculated for the control set increases in parallel with a constantly decreasing error of the training set. As a consequence, completely trained neural networks can reliably forecast the properties of only those compounds which are included in the training set; for other compounds, the quality of forecasting is much worse. This effect is especially pronounced when the first phase of neural network training is either short-lasting or absent. 'Overtraining' can be avoided by reducing the number of adjustable parameters of a neural network model either actual (by decreasing the number of input parameters of the neural network) or effective[86] (by 'regularisation', *e.g.,* Bayesian regularisation[88, 89]).

Early termination of training of neural networks at the moment of phase changes is yet another popular procedure. In the majority of publications, the forecasting ability of neural networks was estimated using the same control set as that used in previous studies for establishing the starting point for 'overtraining'. However, in this case the error can be underestimated (see Ref. 90).

6. Analysis of activities of hidden neurons

An analysis of results obtained with a three-layered neural network used for the simulation of structure–property relationships has shown that hidden neurons can be used for cluster analysis of starting compounds.

Kvasnicka *et al*.[91] have carried out a recurrent cluster analysis aimed at classification of functional groups. Each functional group was represented as an output vector of hidden neurons which was normalised for all compounds under investigation using standard statistical methods.

An analysis of activities of hidden neurons of a three-layered associative neural network aimed at classification of data from the training set was used for the determination of the content of eight fatty acids in samples of olive oil produced in nine different regions of Italy.[31] The use of this approach has made it possible to obtain a more precise, in comparison with the Kohonen neural network, classification of samples with respect to oil-producing regions.

7. Analysis of derived neural network relationships

The need for reliable interpretation of derived models consistent with the data available from the corresponding branches of chemistry, physics or biology is an extremely important aspect which determines the application of virtually all statistical methods including neural networks to the construction of structure–property models in chemistry (particularly those based on the use of physicochemical and quantum-chemical descriptors). Very often, it is this interpretation that represents the most weighty argument in favour of the validity of the constructed model. The impossibility of establishing clear-cut relationships between the properties under study and the input parameters of neural networks, *viz*., the presentation of a neural network as a 'black box', is still the main argument against the application of neural networks to the solution of various chemical problems.

The majority of papers do not provide any interpretation of neural network models; however, the possibility of analysis of neural network relationships has been demonstrated[92, 93] as well as that of estimation of contributions of input parameters to neural network models in neural networks with one output neuron based on analytical formulas for the calculation of partial derivatives of neural network functions.[94–96] It was recommended to supplement a standard sigmoidal transfer function with a parameter b which reflects the degree of mixing of linear and nonlinear functions.

$$\mathrm{Out}_j = \frac{\beta}{(1 + \mathrm{e}^{-\alpha \mathrm{Net}_j})} + (1 - b)\mathrm{Net}_j,$$

where Out_j is the output signal of the neuron j; Net_j is the net input signal of the neuron j; α and β are the coefficients.

The use of specially computed statistical parameters which characterise the function of a trained neural network has been proposed.[97] These parameters allow interpretation of neural network models by conventional methods similar to those used for the interpretation of linear regression models. Indeed, the significance of one or another parameter of a linear regression equation is determined by the value and sign of the corresponding numerical coefficient which, in turn, is equal to the value of a partial derivative of the regression function with respect to this particular parameter. The coefficients are calculated on the whole set of experimental data in the course of regression analysis. A similar interpretation of neural network models demands computation of partial derivatives of properties under study on all examples in succession from the training set for all descriptors. This approach enables computation of a set of statistical functions which allows estimation of the contribution of all original descriptors to the derived neural network models.

8. Choice of a function for estimating computational errors

The derived neural network models are usually estimated on the basis of standard statistical functions by calculation of, *e.g*., root-mean-square errors,[61, 98] correlation coefficients,[36, 99] recognition coefficients,[43, 100] *etc*.

The forecasting ability of neural network models is often estimated using a cross-validation procedure (see, *e.g*., Refs 35, 82, 101).

9. Comparison of a neural network algorithm with standard statistical methods used for information processing

In many cases, conclusions on the adequacy of the use of neural network simulation are drawn by comparing models obtained using neural network algorithms with the models obtained by other methods.

Thus the results of neural network classification of compounds are compared with the results of the hierarchical cluster analysis, analysis of the main components,[32] the results of PLS (Partial Least Squares) in the framework of CoMFA (Comparative Molecular Field Analysis),[9] chemical expert systems,[56] *etc*.

Neural network models of structure–property relationships are most frequently compared with equations obtained by the multiple linear regression method.[77, 102, 103] Most authors conclude that the use of neural network methods ensures the best results in both the description of structure–property relationships and forecasting properties of novel compounds. However, a conclusion[104] about a poor forecasting ability of neural networks in comparison with multiple linear regression can be explained by an incorrect choice of parameters for a neural network model.

Some papers[95, 105] point to a relationship between neural network simulation and multiple regression analysis. It was shown that the results obtained with neural networks containing one hidden layer and utilising a linear transfer function for all

Table 1. Characteristics of neural network models obtained in the simulation of relationships structure – spectral properties and structure – retention index in chromatography.

The property that is simulated	Compounds under study	Number of compounds	Descriptors	Architecture of neural networks		Training results	Statistical characteristics of neural network training [a]	Ref.
				number of neurons	type of neural network [b]			
IR spectra	monosubstituted benzenes	185	topological	32 – (10 × 10) – 128	A	for different compounds	$R = 0.42 – 0.94$	99
	aliphatic compounds containing 4 – 100 carbon atoms	200	calculated IR spectra (PM3)	254 – (from 76 to 650) – 1	B	decrement of errors of the PM3 method for different compounds (from 10% to ~1.25%)	–	108
^{13}C NMR spectra	acyclic alkanes	21	fragmental	12 – 2 – 4 – 6 – 2 – 1	C	calculated values of chemical shifts are close to experimental values	–	24
	monosubstituted benzenes	20	topological and physicochemical	11 – 6 – 4	D	satisfactory classification and forecasting of chemical shifts	–	22
	the same	20	the same	11 – 6 – 4	E	the same	–	22
	heterogeneous set	40 000	fragmental	360 – (5, 10 or 20) – 1	D	averaged results of training of 3 neural networks	$MD_t = 1.79$, $SD_t = 2.10$, $R_t = 0.979$; $MD_p = 1.75$, $SD_p = 1.97$, $R_p = 0.981$ (ppm)	109
	pyranoses and pyranosides	55	topological, electronic, geometrical	11 – 5 – 1	D	–	$RMS_t = 1.03$, $RMS_{cv} = 0.766$, $RMS_p = 1.11$ (ppm)	110
	furanoses and furanosides	56	the same	8 – 5 – 1	D	–	$RMS_t = 1.58$, $RMS_{cv} = 0.995$, $RMS_p = 0.898$ (ppm)	110
	ribonucleosides	17	quantum-chemical	4 – 2 – 1	D	–	$RMS_t = 0.69$, $RMS_{cv} = 0.47$, $RMS_p = 0.39$ (ppm)	111
	acyclic alkenes	130	topological	12 – 4 – 1	D	four networks with different transfer functions	$SD_t = 0.59 – 0.63$, $SD_p = 0.89 – 1.07$ (ppm)	112
NMR spectra	amino acids	123	physicochemical	hierarchical networks: 52 – 4 – 3, 39 – 4 – 3	F	–	$C_t = 77\%$ for $^{13}C_\alpha$ $C_t = 75.3\%$ for $^1H_\alpha$ $C_t = 78.3\%$ for ^{1}HN	44
Retention indices in gas chromatography	alkanes, alkenes, alcohols, esters, ketones	216	types of bonds	29 – 2 – 1	D	–	$RMS_t = 14.7$, $RMS_p = 23.6$	102
Retention indices in reversed-phase chromatography	steroids	85	quantum-chemical	11 – 3 – 1	D	–	$SD_p = 0.08$, $SD_{cv} = 0.09$	113

Table 1 (continued).

The property that is simulated	Compounds under study	Number of compounds	Descriptors	Architecture of neural networks		Training results	Statistical characteristics of neural network training [a]	Ref.
				number of neurons	type of neural network [b]			
Kovats indices	substituted phenols	43	molar refraction, dipole moment	2–3–6–1	A	–	$RMS_p = 1.1$ for the support SE-30 $RMS_p = 2.1$ for the support OV-225 $RMS_p = 0.8$ for the support NGA	37
Retention indices in thin-layer chromatography	substituted benzoic acids	22	quantum-chemical	7–4–1	D	–	$R = 0.973$	103

Note. Here and in Tables 2–4, the architectures of neural networks are described by the number of neurons for each neural network layer starting from the input neuron and ending with the output neuron.

[a] The parameters for assessing neural network training results: *R* is the correlation coefficient; *MD* is the mean deviation; *RE* is the relative error; *SD* is the standard deviation; *RMS*, *MSE* and *S* are the root-mean-square errors; *C* is the recognition coefficient; *SE* is the standard error; *MAE*, *AAE* are the mean absolute errors; *RSD* is the relative standard deviation; *ME* is the mean forecasting error; *ADD* is the standard absolute deviation; *RV* is the residual variation and *SSO* is the sum of squared differences between the calculated and experimental values. The letterings for the subscript indices are as follows: t is the training set; p and v are the control sets; o designates 'for all compounds' and cv is the cross-validation.

[b] A is the back propagation neural network; B is the feedforward neural network; C is the neural network reproducing the topology of molecules under study; D is the feedforward neural network trained according to the delta rule; E is the recurrent neural network; F is the feedforward neural network trained by the resilient propagation method; G is the Kohonen neural network; H is the wavelet neural network; I is the feedforward neural network trained by the Broyden–Fletcher–Goldfarb–Shanno method; J is the radical-basic neural network; K is the fuzzy ARTMAP neural network; L is the feedforward neural network with additional direct bonds between input and output layer neurons trained by the delta rule; M is the feedforward neural network with additional direct bonds between input and output layer neurons trained by the Broyden–Fletcher–Goldfarb–Shanno method; N is the feedforward neural network trained by the scaled conjugated gradient method; O is the feedforward neural network trained with the help of Bayesian regularisation; P is the Kohonen nonlinear distribution neural network; Q is the feedforward functional links neural network.

Table 2. Characteristics of neural network models obtained in the simulation of structure – reactivity relationships of organic compounds.

The property that is simulated	Compounds under study	Number of compounds	Descriptors	Architecture of neural networks		Training results	Statistical characteristics of neural network training [a]	Ref.
				number of neurons	type of neural network [a]			
Cleavage of the C – C bond between the α- and β-atoms relative to the carbonyl group	carbonyl compounds	32	fragmental	8 – (9 × 9)	G	clear-cut separation of the starting array of compounds into four clusters according to bond cleavage type	–	28
Heterolytic bond cleavage	aliphatic compounds	29	physicochemical	7 – (11 × 11)	G	correct classification of reactions according to bond cleavage type for both the training and control sets	–	2
	the same	29	"	7 – 3 – 1	D			
Ratio of isomers upon electrophilic substitution	monosubstituted benzenes	45	quantum-chemical	6 – 10 – 2	D	–	RMS_t = 5.2%, RMS_p = 19.8%	2, 56
	the same	45	tables of bonds of substituents	25 – 5 – 2	D	–	RMS_t = 0.3%, RMS_p = 12.1%	2, 56
Yield of *m*-isomers upon nitration	"	31	fragmental	9 – 4 – 4 – 1	D	good correlation with experimental data for the training set; for the control set the result is at 'semiquantitative level'	–	23
	"	31	"	9 – 5 – 5 – 1	I			
	"	31	"	9 – 6 – 6 – 1	I			
The Markownikoff addition of hydrogen halides	alkenes	25	connectivity matrix	36 – 8 – 8 – 28	D	correct recognition of the addition site for the training and control sets	–	57
Direct and retro Diels – Alder reactions	cycloalkenes	36	the same	120 – 36 – 36 – 105	D	correct classification into forward and retro reactions and identification of the main product of the forward reaction for training and control sets	–	57
Elimination of hydrogen halides according to the Zaitsev rule	alkenes	83	"	66 – 24 – 24 – 55	D	correct identification of the main reaction product for the training and control sets	–	57
Constants of complex formation with α-cyclodextrin	substituted benzenes	24	physicochemical	3 – 3 – 3 – 3 – 1	D	–	R = 0.97, SD_t = 0.22 ($\ln K_a$)	114
Constants of complex formation with β-cyclodextrin	mono- and 1,4-disubstituted benzenes	40	"	6 – 6 – 1	H	–	R = 0.992, SD_t = 0.089 ($\ln K_a$)	115
Heats of formation of aromatic salts	dihydropyridines and related heterocyclic compounds	71		22 – 4 – 1	D	–	R = 0.92, SD_t = 0.298, SD_p = 1.80 kcal mol^{-1}	116

Table 2 (continued).

The property that is simulated	Compounds under study	Number of compounds	Descriptors	Architecture of neural networks		Training results	Statistical characteristics of neural network training [a]	Ref.
				number of neurons	type of neural network [a]			
Addition rate constants of the methyl radical	unsaturated compounds of various types	191	quantum-chemical	6–4–1	D	–	$RMS_t = 0.381$, $RMS_p = 0.496$	53
	the same	191	"	7–3–1	D	–	$RMS_t = 0.424$, $RMS_p = 0.409$ ($\log k$)	53
Reaction with tetrazolium blue, composition of a multicomponent mixture	cortisone and hydrocortisone mixtures	40	"	6–4–2	D	–	$SE_t = 3.47\%$, $SE_p = 4.12\%$ for cortisone; $SE_t = 5.12\%$, $SE_p = 5.25\%$ for hydrocortisone	117

[a] Designations as in Table 1.

Table 3. Characteristics of neural network models obtained in the simulation of structure – physicochemical property relationships of organic compounds.

The property that is simulated	Compounds under study	Number of compounds	Descriptors	Architecture of neural networks		Statistical characteristics of neural network training [a]	Ref.
				number of neurons	type of neural network [a]		
Boiling temperature	alkyl halides $C_1 - C_4$	171	topological	7 – 14 – 1	D	$R = 0.995$	118
	acyclic ethers, peroxides, acetals and their sulfur analogues	185	fragmental	20 – 5 – 1	D	$R = 0.998$, $SE_t = 2.9$ °C, $SE_p = 5.1$ °C	82
	the same	185	topological	3 – 20 – 1	J	$R = 0.990$, $SE_t = 4.9$ °C, $SE_p = 5.9$ °C	39
	tetrahydrofurans, thiophenes, furans, pyrans	299	topological and electronic	16 – 3 – 1	D	$RMS_{cv} = 8.49$ °C	119
	pyridines	291	the same	7 – 3 – 1	D	$RMS_{cv} = 15.8$ °C	119
	heterogeneous set	298	electrotopological	19 – 5 – 1	D	$R = 0.9975$, $MAE_t = 3.86$ °C, $MAE_p = 4.57$ °C	120
	fluoro-substituted ethanes and propanes	31	topological	8 – 6 – 1	D	$MAE_t = 4.7$ °C	121
	alkenes	82	"	5 – 5 – 1	D	$RSD_t = 4.88\%$, $RSD_{cv} = 3.5\%$	122
	alkanes	140	topological and physicochemical	7 – 4 – 1	D	$AAE_t = 1.65$ °C, $AAE_p = 1.73$ °C, $AAE_o = 1.54$ °C	42
	"	140	the same	7 – 4 – 1	K	$AAE_t = 0.81$ °C, $AAE_p = 1.30$ °C	42
	alkenes	144	"	7 – 10 – 1	D	$AAE_t = 6.79$ °C, $AAE_p = 6.45$ °C, $AAE_o = 4.42$ °C	42
	"	144	"	7 – 10 – 1	K	$AAE_t = 0.73$ °C, $AAE_o = 0.95$ °C	42
	alkanes, alkenes and alkynes	327	"	7 – 9 – 1	D	$AAE_t = 6.09$ °C, $AAE_p = 4.68$ °C, $AAE_o = 4.85$ °C	42
	the same	327	"	7 – 9 – 1	K	$AAE_t = 1.15$ °C, $AAE_o = 1.35$ °C	42
	heterogeneous set	298	topological, electronic and mixed	8 – 3 – 1	I	$RMS_t = 7.75$ °C, $RMS_{cv} = 7.17$ °C, $RMS_p = 8.69$ °C	123
	the same	400	topological	26 – 36 – 2	J	$AAE_t = 11$ °C, $AAE_p = 14$ °C	40
	hydrocarbons	134	"	7 – 8 – 6	D	$ME_p = 1.19\%$	101
	heterogeneous set	421	the presence of functional groups	36 – 3 – 4	L	$ADD_p = 2.9\%$	19
Critical temperature	the same	165	electrotopological	19 – 4 – 1	D	$R = 0.9965$, $MAE_t = 4.39$ °C, $MAE_p = 5.59$ °C	120
	fluoro-substituted ethanes and propanes	38	topological and boiling temperature	9 – 4 – 1	D	$MAE_t = 5.9$ °C	121
	heterogeneous set	421	the presence of functional groups	36 – 3 – 4	L	$ADD_p = 3.1\%$	19
	alkanes	69	topological	12 – 5 – 6	D	$R = 0.994$, $S_t = 3.80$, $S_v = 3.94$	124
	"	69	fragmental	14 – 6 – 6	D	$R = 0.995$, $S_t = 3.37$, $S_v = 3.58$	124
Self-ignition temperature	acyclic hydrocarbons with low self-ignition temperature	47	topological and quantum-chemical	5 – 3 – 1	I	$RMS_t = 8.77$ °C, $RMS_{cv} = 6.88$ °C, $RMS_p = 5.11$ °C	125
	acyclic hydrocarbons with high self-ignition temperature	51	the same	6 – 2 – 1	I	$RMS_t = 18.5$ °C, $RMS_{cv} = 17.0$ °C, $RMS_p = 15.7$ °C	125

Table 3 (continued).

The property that is simulated	Compounds under study	Number of compounds	Descriptors	Architecture of neural networks		Statistical characteristics of neural network training [a]	Ref.
				number of neurons	type of neural network [a]		
Self-ignition temperature	nitrogen-containing organic compounds	40	topological and quantum-chemical	6–2–1	I	RMS_t = 34.9 °C, RMS_{cv} = 23.6 °C, RMS_p = 28.2 °C	125
	organic compounds containing oxygen or sulfur atoms	132	the same	7–5–1	I	RMS_t = 30.8 °C, RMS_{cv} = 29.7 °C, RMS_p = 32.5 °C	125
Ignition temperature	heterogeneous set	400	topological	26–36–2	J	AAE_t = 10 °C, AAE_p = 12 °C	40
Heat of evaporation	fluoro-substituted ethanes and propanes	38	"	8–4–1	D	MAE_t = 1.1 kJ mol^{-1}	121
	alkanes	69	"	12–5–6	D	R = 0.994, S_t = 0.44, S_v = 0.51	124
	"	69	fragmental	14–6–6	D	R = 0.996, S_t = 0.44, S_v = 0.56	124
Thermal capacity	hydrocarbons	134	topological	7–8–6	D	ME_p = 0.87%	101
Density	"	134	"	7–8–6	D	ME_p = 0.60%	101
	fluoro-substituted ethanes and propanes	38	"	8–4–1	D	MAE_t = 0.03 g cm^{-3}	121
	alkenes	82	"	5–5–1	D	RSD_t = 0.43%, RSD_{cv} = 0.4%	122
Refraction index	hydrocarbons	134	"	7–8–6	D	ME_p = 0.19%	101
	alkenes	82	"	5–5–1	D	RSD_t = 0.13%, RSD_{cv} = 0.14%	122
The Gibbs energy	hydrocarbons	134	"	7–8–6	D	ME_p = 1.36%	101
Enthalpy	"	134	"	7–8–6	D	ME_p = 1.42%	101
Critical volume	heterogeneous set	421	the presence of functional groups	36–3–4	L	ADD_p = 3.4%	19
Acentric factor	the same	421	the same	36–3–4	L	ADD_p = 9.1%	19
Molar volume	alkanes	69	topological	12–5–6	D	R = 0.999, S_t = 0.84, S_v = 0.89	124
	"	69	fragmental	14–6–6	D	R = 0.999, S_t = 0.88, S_v = 1.10	124
Molar refraction	"	69	topological	12–5–6	D	R = 1.000, S_t = 0.15, S_v = 0.18	124
	"	69	fragmental	14–6–6	D	R = 0.999, S_t = 0.20, S_v = 0.18	124
Critical pressure	"	69	topological	12–5–6	D	R = 0.984, S_t = 0.46, S_v = 0.39	124
	"	69	fragmental	14–6–6	D	R = 0.986, S_t = 0.44, S_v = 0.23	124
	heterogeneous set	421	boiling temperature, critical temperature, critical volume	3–3–1	L	ADD_p = 6.4%	19
Surface tension	alkanes	69	topological	12–5–6	D	R = 0.996, S_t = 0.18, S_v = 0.28	124
	"	69	fragmental	14–6–6	D	R = 0.996, S_t = 0.17, S_v = 0.17	124

Table 3 (continued).

The property that is simulated	Compounds under study	Number of compounds	Descriptors	Architecture of neural networks		Statistical characteristics of neural network training [a]	Ref.
				number of neurons	type of neural network [a]		
Saturated vapour pressure [log *VP* (Pa)]	compounds with heteroatoms	420	topological	8–3–1	I	$RMS_t = 0.26$, $RMS_p = 0.37$, $RMS_{cv} = 0.29$	54
	the same	420	"	10–4–1	I	$RMS_t = 0.19$, $RMS_p = 0.33$, $RMS_{cv} = 0.24$	54
	hydrocarbons and halogenohydrocarbons	352	topological, geometrical and electronic	7–3–1	I	$RMS_t = 0.163$, $RMS_{cv} = 0.163$, $RMS_p = 0.209$	126
Viscosity [log *M* (mPa s)]	heterogeneous set	361	experimental physico-chemical values	9–3–1	D	$R = 0.958$, $RMS_t = 0.118$, $RMS_p = 0.161$	127
Solubility in water [log *S* (mol litre^{-1})]	the same	350	quantum-chemical	17–18–1	D	$SD_t = 0.23$, $SD_p = 0.43$	128
	"	332	topological, electronic and geometrical	9–6–1	I	$RMS_t = 0.394$, $RMS_{cv} = 0.358$, $RMS_p = 0.343$	129
	heterogeneous set of drugs	211	electrotopological and topological	23–5–1	D	$R_t^2 = 0.90$, $SE_t = 0.46$, $R_p^2 = 0.86$, $SE_p = 0.53$	130
	heterogeneous set	136	topological, electronic, geometrical and mixed	9–3–1	I	$RMS_t = 0.145$, $RMS_{cv} = 0.151$, $RMS_p = 0.166$	131
Solubility in carbon dioxide in a supercritical state [log *S* (mol litre^{-1})]	the same	58	topological and quantum-chemical	7–2–1	I	$RMS_t = 0.65$, $RMS_{cv} = 0.68$, $RMS_p = 0.64$	132
Activity coefficient at infinite dilution ($\log \gamma^{\infty}$)	"	325	the same	12–6–1	I	$RMS_t = 0.472$, $RMS_{cv} = 0.538$, $RMS_p = 0.484$	133
Lipophilicity (log *P*) [b]	"	250	quantum-chemical	8–5–1	D	$R = 0.923$, $SE_p = 0.379$	134
	"	250	"	13–6–1	D	$R = 0.952$, $SE_p = 0.300$	134
	"	1870	molecular weight and electrotopological	39–5–1	D	$R_t^2 = 0.90$, $RMS_t = 0.46$, $R_p^2 = 0.94$, $RMS_p = 0.4$	135
	"	323	quantum-chemical	6–7–1	I	$SE_p = 0.30$	136
	"	323	"	11–3–1	D	$SD_t = 0.31$, $SD_p = 0.29$, $SD_{cv} = 0.32$	113
	compounds with heteroatoms	7719	topologo-physicochemical	35–32–1	D	$R_t = 0.97$, $RMS_t = 0.37$, $R_p = 0.98$, $RMS_p = 0.39$	137
Phase-transition temperature (°C)	nematic liquid-crystalline compounds	17383	fragmental	205–100–1	D	$SD_t = 3.8$, $SD_p = 16.4$	138, 139
	the same	6304	"	205–10–1	D	$SD_p = 18.8$	138, 139
Solvatochromic polarity–polarisability index	heterogeneous set	333	topological, electrostatic, quantum-chemical, structural	16–7–1	D	$R^2 = 0.980$	77
Molar adsorption ($\log \varepsilon$)	asymmetric phosphobisazo derivatives	43	topological	4–4–3–1	I	$RMS_t = 0.05669$, $RMS_p = 0.09621$	61

Table 3 (continued).

The property that is simulated	Compounds under study	Number of compounds	Descriptors	Architecture of neural networks		Statistical characteristics of neural network training [a]	Ref.
				number of neurons	type of neural network [a]		
Impact sensitivity ($\log H_{50\%}$) [c]	nitro compounds	204	topological, quantum-chemical	11 – 2 – 1	D	$R = 0.941$, $SE_p = 0.154$	62
	"	204	the same	13 – 2 – 1	D	$R = 0.937$, $SE_p = 0.159$	62
	"	204	"	13 – 3 – 1	D	$R = 0.95$, $SE_p = 0.13$	140
Position of the long-wave absorption band	symmetrical cyanine dyes	398	quantum-chemical, indicators of the presence of substituents	8 – 10 – 1	D	$R_s = 0.9928$, $S_t = 10.6$ nm, $S_v = 7.0$ nm	141
Interatomic distances	compounds containing two heteroatoms	2615	geometrical	78 – 8 – 1	D	$C_t = 80\%$, $C_p = 80\%$ for remote atoms; $C_t = 66\%$ for adjacent atoms	100
Odour	aliphatic alcohols	99	topological	6 – 3 – 1	D	$C_t = 100\%$, $C_p = 85\%$	142
Total energy of π-electrons	unsubstituted polycyclic hydrocarbons	265	"	3 – 3 – 1	D	$SD_t = 0.06$, $SD_{cv} = 0.27$	143
Inductive and resonance constants of substituents	monovalent functional groups [d]	37	"	14 – 8 – 2	D	–	144

[a] Designations as in Table 1.

[b] $\log P$ is the common logarithm of the distribution coeficient in the n-octanol – water system.

[c] $H_{50\%}$ is the height for which the explosion occurs with 50% probability.

[d] For compounds under study, the result of neural network training is a correct classification of substituents and a good estimation of the values of substituent constants for both training and control sets.

Table 4. Some characteristics of neural network models obtained in the simulation of structure – biological activity relationships of organic compounds.

The property that is simulated	Number of compounds	Compounds under study	Descriptors	Architecture of neural networks		Training results	Statistical characteristics of neural network training [a]	Ref.
				number of neurons	type of neural network [a]			
Affinity to transport proteins (globulins)	31	corticosteroids and testosterones	physicochemical	32 – (10 × 10)	G	clear-cut separation of steroids into three clusters (strong, medium and weak affinity)	–	99
		the same	map of atomic cooordinates	9 – (7 × 7)	G	separation of steroids into two clusters (strong and weak affinity)	–	29
		"	map of atomic coordinates and charges	9 – (7 × 7)	G	clear-cut separation according to the type of receptors for training and control sets	–	30
Affinity to estrogen receptors	22	disubstituted hexestrol derivatives	topological and physicochemical	2 – 6 – 1	D	–	$RMS_t = 0.09131$ (log *RBA*) [b]	145
	16	hexestrol ethers	topological	1 – 4 – 2 – 1	D	–	$RMS_t = 0.018219$ (log *RBA*)	145
	16	deoxyhexestrols	"	1 – 4 – 2 – 1	D	–	$RMS_t = 0.018219$ (log *RBA*)	145
Affinity to progesterone receptors	55	androst-4-en-3-one derivatives	physicochemical	6 – 2 – 1	M	–	$R = 0.96$, $RMS_t = 0.30$, $RMS_p = 1.32$ (log IC_{50}) [c]	20
Affinity to benzodiazepine receptors	57	1,4-benzodiazepin-3-ones	"	10 – 3 – 1	N	–	$R_t = 0.954$, $SD_t = 0.007$, $R_{cv} = 0.901$ (log IC_{50})	146
$GABA_A$				6 – 2 – 1	N	–	$SD_{cv} = 0.009$, $R_{cv} = 0.938$	146
Inhibition of reverse transcriptase of the human immunodeficiency virus	44	derivatives of AZT [d], TIBO [e] and related compounds	topological	4 – 3 – 2	D	–	$C_t = 100\%$, $C_p = 86\%$	74, 75
	107	1-[(2-hydroxyethoxy)-methyl]-6-(phenylthio)-thymine derivatives	quantum-chemical	6 – 6 – 1	D	–	$MSE_t = 0.073$, $MSE_p = 0.372$ (log $1/C$) [f]	147
Inhibition of protein tyrokinase $p56^{lck}$	105	substituted flavonoids	physicochemical constants of substituents	6 – (10 × 10) – 1	A	–	$R^2 = 0.92$, $SD_p = 0.55$ (log $1/IC_{50}$)	35
			quantum-chemical	3 – (10 × 10) – 1	A	–	$R^2 = 0.96$, $SD_p = 0.40$ (log $1/IC_{50}$)	35
Inhibition of 5-lipoxygenase *in vitro*	68	arylhydroxamic acids	physicochemical	1 – 4 – 1	D	–	$RMS_t = 0.09043$ (log $1/K$) [g]	145
Inhibition of dihydrofolate reductase	157	3,4-diamino-6,6-dimethyl-5-phenyldihydrotriazines	"	6 – 4 – 1	D	–	$R = 0.922$, $SD_t = 0.374$ (log IC_{50})	80

Table 4 (continued).

The property that is simulated	Number of compounds	Compounds under study	Descriptors	Architecture of neural networks		Training results	Statistical characteristics of neural network training [a]	Ref.
				number of neurons	type of neural network [a]			
Inhibition of dihydrofolate reductase	61	symmetrical 1-aryl-4,6-diamino-2,2-dimethyl-1,2-dihydrotriazines	physicochemical	3 – 1 – 1	M	–	$R = 0.87$, $SD_t = 0.27$	20
	68	2,4-diamino-(5*R*)-benzylpyrimidines	"	4 – 7 – 1	D	–	$R = 0.97$, $RV_t = 0.0126$, $RV_p = 0.323$	83
Inhibition of acylcholesterol-*O*-acyl transferase	157	aminosulfonyl-, hydroxysulfonyl- and (2,6-diisopropylphenoxy)carbonates, aminosulfonyl- and hydroxysulfonylureas	topological, geometrical and electronic	8 – 3 – 1	I	–	$RMS_t = 0.226$, $RMS_{cv} = 0.208$, $RMS_p = 0.242$ (log IC_{50})	55
Inhibition of [^{3}H]diazepam binding	245	benzodiazepines	topological	50 – 4 – 1	O	–	$R^2 = 0.63$, $SE_p = 0.14$ (log IC_{50})	88, 89
Antifungal activity	103	(3*R*)-1-(3,5-dichlorophenyl)-pyrrolidine-2,5-diones	physicochemical parameters of substituents	5 – (5 × 5)	O	clear-cut separation of starting compounds into groups	–	78
Antihelminthic activity	31	2-hydroxyphenylamides	physicochemical	3 – 3 – 1	D	–	$R = 0.919$, $RMS_t = 0.322$	64
		"	"	53 – (8 × 8)	G	clear-cut separation into groups possessing different activities	–	32
Hypotensive activity	24	(3*R*)-phenylthiopropyl-substituted heterocycles	"	6 – 3 – 1	D	–	$SD_{cv} = 11.6\%$	148
	29	arylacryloylpiperazines	"	7 – 14 – 4	D	–	$C_t = 19/21$, $C_p = 6/8$	149
Anticonvulsant activity	60	benzodiazepine derivatives	"	14 – 15 – 57 – 1	A	–	$RMS_t = 0.16$	38
			"	14 – 28 – 1	D	–	$SD_t = 0.33$	10, 38
			"	5 – 6 – 1	D	–	$R = 0.887$	2, 43
			"	5 – 1	Q	–	$R = 0.876$, $RMS_t = 0.369$	43
Antitumour activity (minimum effective dose for multiple injections)	39	2,5-bis(1-aziridinyl)-*p*-benzoquinones	"	6 – 7 – 37 – 1	A	–	$RMS_t = 0.08$	38
			"	12 – 13 – 37 – 1	A	–	$RMS_t = 0.06$	38
			"	7 – 12 – 1	D	–	$RMS_t = 0.21$	2, 10, 38
			"	13 – 26 – 1	D	–	$RMS_t = 0.21$	10, 38
Antitumour activity (minimum effective dose for single injection)	39		"	6 – 7 – 37 – 1	A	–	$RMS_t = 0.04$	38
			"	12 – 13 – 37 – 1	A	–	$RMS_t = 0.20$	38
			"	7 – 12 – 1	D	–	$RMS_t = 0.25$	2, 10, 38
			"	13 – 26 – 1	D	–	$RMS_t = 0.24$	2, 10, 38
Antitumour activity (optimum dose for multiple injections)	39		"	6 – 7 – 37 – 1	A	–	$RMS_t = 0.02$	38
			"	12 – 13 – 37 – 1	A	–	$RMS_t = 0.05$	38
			"	7 – 12 – 1	D	–	$RMS_t = 0.19$	2, 10, 38

Table 4 (continued).

The property that is simulated	Number of compounds	Compounds under study	Descriptors	Architecture of neural networks		Training results	Statistical characteristics of neural network training [a]	Ref.
				number of neurons	type of neural network [a]			
Antitumour activity (optimum dose for multiple injections)	39		physicochemical	13–26–1	D	–	$RMS_t = 0.18$	2, 10, 38
			"	4–13–1	D	–	$R = 0.960$	43
			"	4–1	Q	–	$R = 0.951$, $RMS_t = 0.169$	43
Antitumour activity (optimum dose for single injection)	39		"	6–7–37–1	A	–	$RMS_t = 0.08$	38
			"	12–13–37–1	A	–	$RMS_t = 0.05$	38
			"	7–12–1	D	–	$RMS_t = 0.20$	2, 10, 38
			"	13–26–1	D	–	$RMS_t = 0.17$	2, 10, 38
Inhibition of methotrexate-sensitive tumour cells	61	3-substituted triazines	"	3–3–1	M	–	$R = 0.86$, $SD_t = 0.25$	20
Inhibition of methotrexate-resistant tumour cells	62		"	3–1	M	–	$R = 0.58$, $SD_t = 0.53$	20
Mechanisms of antitumour activity	141	drugs	activities with respect to different tumour cells	60–7–6	D	–	$C_p = 129/141$	150
Toxicity	91	benzothiazolium salts	fragmental	30–2–1	D	–	$SSO_t = 0.017$, $SSO_p = 0.850$	151
Anticarcinogenic activity	16	carboquinones	physicochemical	7–12–5	D	–	$C_t = 100\%$, $C_p = 3/5$	84, 94, 149
				7–4–5	D	–	$C_t = 100\%$, $C_p = 3/5$	84
Carcinogenic activity	45	mono- and polysubstituted benzenes	topological	90–100–100–1	A	–	$R_t = 0.95$, $R_p = 0.70$	36
			topological and quantum-chemical	90–100–100–1	A	–	$R_t = 0.98$, $R_p = 0.72$	36
			quantum-chemical	90–100–100–1	A	–	$R_t = 1.00$, $R_p = 0.63$	36
	11	polycyclic aromatic hydrocarbons	^{13}C NMR spectra	10–3–1	D	clear-cut separation into active and inactive compounds	–	76
	81	aromatic compounds with nitrogen-containing substituents	quantum-chemical	6–12–2	D	–	$C_t = 84.6\%$, $C_p \approx 80\%$	152
	104		topological, physicochemical and quantum-chemical	12–4–1	D	–	$R^2_{cv} = 0.691$, $MSE_{cv} = 0.0416$ [log $(MW \cdot 1000/ID_{50})$] [h, i]	153
Mutagenic activity	197	aromatic and heteroaromatic compounds	physicochemical	2–4–2–1	D	–	$RMS_t = 0.14331$ (log TA_{98}) [j]	145
			"	4–8–4–1	D	–	$RMS_t = 0.11515$, $RMS_p = 0.1216$	145
			"	4–4–1	D	–	$R_t = 0.919$, $SD_t = 0.789$, $R_p = 0.853$, $SD_p = 1.049$ (log TA_{98})	154

Table 4 (continued).

The property that is simulated	Number of compounds	Compounds under study	Descriptors	Architecture of neural networks		Training results	Statistical characteristics of neural network training [a]	Ref.
				number of neurons	type of neural network [a]			
Mutagenic activity	488	heterogeneous set	fragmental	100 – 10 – 2	D	good separation into clusters according to size and activity	–	69, 155
	54	heterocyclic analogues of pyrene and phenanthrene, fluorenes, biphenyls	energy of LUMO, lipophilicity, presence of a nitro group in the *p*-position	3 – 2 – 1	D	–	$R = 0.87$, $S_t = 1.03$, $S_v = 1.47$ ($\ln N_{his+}$) [k]	156
Inhibition of intracellular ion exchange Na^+/H^+	113	benzoylguanidine derivatives	topological, geometrical, electronic and combined	5 – 4 – 1	D	–	$R_t^2 = 0.812$, $RMS_t = 0.278$, $R_p^2 = 0.697$, $RMS_p = 0.362$ ($\log IC_{50}$)	157
Hallucinogenic activity	64	phenylalkylamines	fragmental	15 – (33 × 33)	G	clear-cut separation of compounds into clusters according to activity	–	28
	35	"	constants of substituents, indicators of the presence of substituents	7 – 2 – 1	D	–	$R = 0.932$, $S_t = 0.55$, $S_v = 0.47$ ($\ln MU$) [l, m]	58
							$R = 0.820$, $S_t = 0.89$, $S_v = 0.54$ ($\ln MU$) [n]	58
Blocking of Ca^{2+} channels	46	1,4-dihydropyridines	constants of substituents	5 – 2 – 1	D	–	$R = 0.832$, $S_t = 0.79$, $S_v = 0.71$ ($\log 1/EC_{50}$) [m, o]	58
					D	–	$R = 0.870$, $S_t = 0.70$, $S_v = 1.59$ ($\log 1/EC_{50}$) [n]	58
Adsorption by human gastric mucosa	86	drugs	quantum-chemical and topological	6 – 4 – 1	D	–	$RMS_t = 9.4\%$, $RMS_{cv} = 19.7\%$, $RMS_p = 16.0\%$	158
Antibacterial activity [p]	111	fluoroquinolones	topological	62 – 2 – 1	D	–		
$MIC \leqslant 0.05$							$C_t = 100\%$, $C_p = 93.88\%$	159
$MIC \leqslant 0.10$							$C_t = 97.28\%$, $C_p = 82.36\%$	
$MIC \leqslant 0.20$							$C_t = 99.53\%$, $C_p = 87.99\%$	

[a] Designations as in Table 1; [b] RBA is the affinity for the receptor; [c] IC_{50} is the concentration causing 50% inhibition; [d] AZT is azidothymidine; [e] TIBO is 4,5,6,7-tetrahydroimidazo[4,5,1-*j*,*k*][1,4]benzodiazepin-2(1*H*)-one; [f] C is the concentration of a compound causing a certain biological response; [g] K is the inhibition constant; [h] MW is the molecular weight; [i] ID_{50} is the dose causing 50% inhibition; [j] TA_{98} is the number of revertants in the Ames test TA_{98}; [k] $N_{his}{}^+$ is the number of revertants in the Ames test TA_{1538}; [l] MU are the muscarine units; [m] the initial set of compounds was supplemented with topologically equivalent structures; [n] without multiplication of the initial set of compounds; [o] EC_{50} is the effective concentration; [p] MIC is the minimum inhibiting concentration /mg ml^{-1}.

neurons are identical with the results obtained by multiple linear regression analysis.

If the neurons of the hidden layer of a three-layered neural network are described by a nonlinear (most commonly, sigmoidal) transfer function and the neuron of the output layer is described by a linear function, the resulting neural network (termed MR neural network) will function similarly to a nonlinear multiple regression.[96, 106, 107]

IV. The use of neural network algorithms for elucidating structure – property relationships

The neural network approach can successfully be employed for elucidating both qualitative and quantitative structure – property relationships. Some brief information about the use of neural network models for the simulation of spectral characteristics, reactivities, physicochemical properties and biological activities of organic compounds is given in Tables 1 – 4.

The information contained in the IR[99, 108] and NMR spectra[22, 24, 44, 109 – 112] as well as chromatographic retention indices[37, 102, 103, 113] is closely related to the structures of organic molecules. In the studies cited in Table 1, the neural network algorithm for processing of such information was successfully used for the solution of miscellaneous practical problems related to the search for relationships between molecular structures and spectral properties, classification and forecasting of structural data and refinement of molecular structures on the basis of spectral information.

In one pioneering study, neural network methods were used for the classification of reactions of organic compounds, *viz.*, according to the cleaved bond type,[2, 28] regiochemistry of the addition reaction,[57] isomer ratios,[2, 23, 56] main reaction products,[57] *etc.* (Table 2). In these studies, connectivity matrices of organic compounds were used as input information for neural network models, since they do not require additional calculations at the preceding steps. Neural network methods can be used for quantitative estimation of reactivities of organic compounds and simulation of complex kinetic processes in the course of chemical transformations.

The publications devoted to the analysis of various physicochemical properties of chemical compounds using neural networks are rather numerous; the most popular characteristics of organic compounds include boiling temperature[19, 39, 40, 42, 82, 101, 118 – 123] (predominantly with the use of topological indices), density,[101, 121, 122] solubility in water,[128 – 131] lipophilicity[134 – 137], *etc.* (Table 3).

Today, simulation of biological activities of organic compounds presents considerable practical interest: the overwhelming majority of publications devoted to the use of artificial neural networks for elucidating structure – property relationships are related to this particular area (Table 4).[31] Some publications describe the results of successful quantitative neural network simulation of affinities of organic compounds to various receptors[20, 145, 146] and transport proteins,[29, 30, 99] simulation of inhibition constants of enzymes,[20, 35, 55, 74, 75, 80, 83, 145, 147] antitumour,[20, 38, 43, 105, 150] carcinogenic,[36, 76, 152, 153] mutagenic[69, 145, 154 – 156] and antibacterial activities,[159] *etc.*

* * *

Artificial neural networks represent a potent comprehensive computational tool for the solution of miscellaneous problems related to processing of chemical information, particularly classification and simulation of disembodied experimental data. The number of publications on the subject is increasing annually, which testifies to the unrelenting interest of chemists in the simulation of structure – property relationships using artificial neural networks.

References

1. J Zupan, J Gasteiger *Anal. Chim. Acta* **248** 1 (1991)
2. J Gasteiger, J Zupan *Angew. Chem., Int. Ed. Engl.* **105** 503 (1993)
3. T Aoyama, H Ichikawa *Chem. Pharm. Bull.* **39** 358 (1991)
4. J A Burns, G M Whitesides *Chem. Rev.* **93** 2583 (1993)
5. J Zupan, J Gasteiger *Neural Networks in Chemistry* (Weinheim: Wiley-VCH, 1999)
6. I V Zaentsev *Neironnye Seti: Osnovnye Modeli* (*Neural Networks: Basic Models)* (Voronezh: Voronezh State University, 1999)
7. M Lehtokangas, J Saarinen *Neurocomputing* **20** 265 (1998)
8. J Y F Yam, T W S Chow *Neurocomputing* **30** 219 (2000)
9. I Kovesdi, M F Dominguez-Rodrigues, L Orfi, G Naray-Szabo, A Varro, J G Papp, P Matyus *Med. Res. Rev.* **19** 249 (1999)
10. J Devillers (Ed.) *Neural Networks in QSAR and Drug Design* (London: Academic Press, 1996)
11. A N Gorban' *Obuchenie Neironnykh Setei* (*Teaching of Neural Networks)* (Moscow: ParaGraf, 1990)
12. D E Rumelhart, G E Hinton, R J Williams, in *Parallel Data Processing* Vol. 1 (Eds D E Rumelhart, J L McClelland) (Cambridge, MA: M.I.T. Press, 1986) p. 318
13. *Stuttgart Neural Network Simulator* User Manual, Version 4.0, 1995
14. *MatLab* User Manual, Version 6.0, 2000
15. T P Vogl, J K Mangis, A K Rigler, W T Zink, D L Alkon *Biol. Cybern.* **59** 256 (1988)
16. L M Patnaik, K Rajan *Neurocomputing* **35** 123 (2000)
17. M Riedmiller, H Braun, in *Proceedings of the IEEE International Conference on Neural Networks, San Francisco, CA, 1993* p. 586
18. J Dennis, R B Schnabel *Numerical Methods for Unconstrained Optimization and Nonlinear Equations* (Englewood Cliffs, NJ: Prentice-Hall, 1983)
19. M J Lee, J T Chen *Ind. Eng. Chem. Res.* **32** 995 (1993)
20. Ajay *J. Med. Chem.* **36** 3565 (1993)
21. L Dolmatova, V Tchistiakov, K Ruckebusch, N Dupuy, J-P Huvenne, P Legrand *J. Chem. Inf. Comput. Sci.* **39** 1027 (1999)
22. V Kvasnicka, S Sklenak, J Pospichal *J. Chem. Inf. Comput. Sci.* **32** 742 (1992)
23. V Kvasnicka, J Pospichal *J. Mol. Struct. (Theochem)* **235** 227 (1991)
24. V Kvasnicka *J. Math. Chem.* **6** 63 (1991)
25. D B Kireev *J. Chem. Inf. Comput. Sci.* **35** 175 (1995)
26. I I Baskin, V A Palyulin, N S Zefirov *Dokl. Akad. Nauk* **333** 176 (1993) [a]
27. I I Baskin, V A Palyulin, N S Zefirov *J. Chem. Inf. Comput. Sci.* **37** 715 (1997)
28. B Bienfait *J. Chem. Inf. Comput. Sci.* **34** 890 (1994)
29. J Polanski *J. Chem. Inf. Comput. Sci.* **36** 694 (1996)
30. J Polanski *J. Chem. Inf. Comput. Sci.* **37** 553 (1997)
31. R Kocjancic, J Zupan *J. Chem. Inf. Comput. Sci.* **37** 985 (1997)
32. V S Rose, I F Croall, H J MacFie *Quantum Struct.-Act. Relat.* **10** 6 (1991)
33. R Hecht-Nielsen *Appl. Opt.* **26** 4979 (1987)
34. R Hecht-Nielsen *Neural Netw.* **1** 131 (1988)
35. M Novic, Z Nikolovska-Coleska, T Solmajer *J. Chem. Inf. Comput. Sci.* **37** 990 (1997)
36. M Vracko *J. Chem. Inf. Comput. Sci.* **37** 1037 (1997)
37. K L Peterson *Anal. Chem.* **64** 379 (1992)
38. K L Peterson *J. Chem. Inf. Comput. Sci.* **35** 896 (1995)
39. H Lohninger *J. Chem. Inf. Comput. Sci.* **33** 736 (1993)
40. J Tetteh, T Suzuki, E Metcalfe, S Howells *J. Chem. Inf. Comput. Sci.* **39** 491 (1999)
41. C H Wan, P de B Harrington *J. Chem. Inf. Comput. Sci.* **39** 1049 (1999)
42. G Espinosa, D Yaffe, Y Cohen, A Arenas, F Giralt *J. Chem. Inf. Comput. Sci.* **40** 859 (2000)
43. Q Liu, S Hirono, I Moriguchi *Chem. Pharm. Bull.* **40** 2962 (1992)
44. W Y Choy, B C Sanctuary, G Zhu *J. Chem. Inf. Comput. Sci.* **37** 1086 (1997)
45. C Cai, P de B Harrington *J. Chem. Inf. Comput. Sci.* **39** 874 (1999)
46. V V Kovalishyn, I V Tetko, A I Luik, V V Kholodovych, A E P Villa, D J Livingstone *J. Chem. Inf. Comput. Sci.* **38** 651 (1998)

47. A Micheli, A Sperduti, A Starita, A M Bianucci *J. Chem. Inf. Comput. Sci.* **41** 202 (2001)
48. R Todeschini, V Consonni *Handbook of Molecular Descriptors* (Weinheim: Wiley-VCH, 2000)
49. N V Artemenko, I I Baskin, V A Palyulin, N S Zefirov *Dokl. Akad. Nauk* **381** 203 (2001) [a]
50. M I Stankevich, I V Stankevich, N S Zefirov *Usp. Khim.* **57** 337 (1988) [*Russ. Chem. Rev.* **57** 191 (1988)]
51. L Pogliani *Chem. Rev.* **100** 3827 (2000)
52. M Karelson, V S Lobanov, A R Katritzky *Chem. Rev.* **96** 1027 (1996)
53. G A Bakken, P C Jurs *J. Chem. Inf. Comput. Sci.* **39** 508 (1999)
54. H E McClelland, P C Jurs *J. Chem. Inf. Comput. Sci.* **40** 967 (2000)
55. S J Patankar, P C Jurs *J. Chem. Inf. Comput. Sci.* **40** 706 (2000)
56. D W Elrod, G M Maggiora, R G Trenary *J. Chem. Inf. Comput. Sci.* **30** 477 (1990)
57. D W Elrod, G M Maggiora, R G Trenary *Tetrahedron Comput. Methodol.* **3** 163 (1990)
58. I I Baskin, N M Halberstam, T V Mukhina, V A Palyulin, N S Zefirov *SAR QSAR Environ. Res.* **12** 401 (2001)
59. V N Viswanadhan, G A Mueller, S C Basak, J N Weinstein *J. Chem. Inf. Comput. Sci.* **41** 505 (2001)
60. A M Kshirsagar *Multivariate Analysis* (New York: Marcel Dekker, 1972)
61. M Guo, L Xu, H Li, C-Y Hu *Anal. Sci.* **12** 291 (1996)
62. H Nefati, J-M Cense, J-J Legendre *J. Chem. Inf. Comput. Sci.* **36** 804 (1996)
63. G Castellano, A M Fanelli *Neurocomputing* **31** 1 (2000)
64. S-S So, M Karplus *J. Med. Chem.* **39** 1521 (1996)
65. S-S So, M Karplus *J. Med. Chem.* **40** 4347 (1997)
66. S-S So, M Karplus *J. Med. Chem.* **40** 4360 (1997)
67. D Rogers, A J Hopfinger *J. Chem. Inf. Comput. Sci.* **34** 854 (1994)
68. B T Luke *J. Chem. Inf. Comput. Sci.* **34** 1279 (1994)
69. M W Brinn, M P Payne, P T Walsh *Chem. Eng. Res. Design* **71** 337 (1993)
70. E D Karnin *IEEE Trans. Neural Netw.* **1** 239 (1990)
71. J H Wikel, E R Dow *Bioorg. Med. Chem. Lett.* **3** 645 (1993)
72. R Reed *IEEE Trans. Neural Netw.* **4** 740 (1993)
73. I V Tetko, A E P Villa, D J Livingstone *J. Chem. Inf. Comput. Sci.* **36** 794 (1996)
74. I V Tetko, V Yu Tanchuk, A I Luik, in *Proceedings of AIP Conference 330 (E.C.C.C. 1 Computational Chemistry) (F.E.C.S. Conference), Nancy, 1995* p. 761
75. I V Tetko, V Yu Tanchuk, N P Chentsova, S V Antonenko, G I Poda, V P Kukhar, A I Luik *J. Med. Chem.* **37** 2520 (1994)
76. Y Isu, U Nagashima, T Aoyama, H Hosoya *J. Chem. Inf. Comput. Sci.* **36** 286 (1996)
77. D Svozil, J G K Sevcik, V Kvasnicka *J. Chem. Inf. Comput. Sci.* **37** 338 (1997)
78. D Domine, J Devillers, D Wienke, L Buydens *Quantum Struct.-Act. Relat.* **15** 395 (1996)
79. D Domine, J Devillers, D Wienke, L Buydens *J. Chem. Inf. Comput. Sci.* **37** 10 (1997)
80. T A Andrea, H Kalayeh *J. Med. Chem.* **34** 2824 (1991)
81. X-H Song, Z Chen, R-Q Yu *Sci. China, Ser. B* **36** 1443 (1993)
82. D Cherqaoui, D Villemin, A Mesbah, J M Cense, V Kvasnicka, in *Proceedings ofAIP Conference 330 (E.C.C.C. 1 Computational Chemistry) (F.E.C.S. Conference), Nancy, 1995* p. 767
83. S-S So, W G Richards *J. Med. Chem.* **35** 3201 (1992)
84. I V Tetko, A I Luik, G I Poda *J. Med. Chem.* **36** 811 (1993)
85. B Curry, D E Rumelhart *Tetrahedron Comput. Methodol.* **3** 213 (1990)
86. A S Weigend, B A Huberman, D E Rumelhart *Int. J. Neural Syst.* **1** 193 (1990)
87. D T Manallack, D J Livingstone *Eur. J. Med. Chem.* **34** 195 (1999)
88. F R Burden, D A Winkler *J. Med. Chem.* **42** 3183 (1999)
89. F R Burden, M G Ford, D C Whitley, D A Winkler *J. Chem. Inf. Comput. Sci.* **40** 1423 (2000)
90. I I Baskin, M I Skvortsova, V A Palyulin, N S Zefirov *Found. Comput. Decis. Sci.* **22** 107 (1997)
91. V Kvasnicka, S Sklenak, J Pospichal *J. Am. Chem. Soc.* **115** 1495 (1993)
92. J Y Choi, C-H Choi *IEEE Trans. Neural Netw.* **3** 101 (1992)
93. S-H Oh, Y Lee *IEEE Trans. Neural Netw.* **6** 1005 (1995)
94. T Aoyama, H Ichikawa *J. Chem. Inf. Comput. Sci.* **32** 492 (1992)
95. T Aoyama, H Ichikawa *Chem. Pharm. Bull.* **39** 372 (1991)
96. H Ichikawa, T Aoyama *SAR QSAR Environ. Res.* **1** 115 (1993)
97. I I Baskin, A O Ait, N M Halberstam, V A Palyulin, N S Zefirov *SAR QSAR Environ. Res.* **13** 35 (2002)
98. G M Maggiora, D W Elrod, R G Trenary *J. Chem. Inf. Comput. Sci.* **32** 732 (1992)
99. J H Schuur, P Selzer, J Gasteiger *J. Chem. Inf. Comput. Sci.* **36** 334 (1996)
100. S N Jordan, A R Leach, J Bradshaw, in *Proceedings of AIP Conference 330 (E.C.C.C. 1 Computational Chemistry) (F.E.C.S. Conference), Nancy, 1995* p. 550
101. A A Gakh, E G Gakh, B G Sumpter, D W Noid *J. Chem. Inf. Comput. Sci.* **34** 832 (1994)
102. A Bruchmann, P Zinn, C M Haffer *Anal. Chim. Acta* **283** 869 (1993)
103. R C Glen, V S Rose, J C Lindon, R J Ruane, I D Wilson, J K Nicholson *J. Planar Chromatogr.* **4** 432 (1991)
104. D T Manallack, D D Ellis, D J Livingstone *J. Med. Chem.* **37** 3758 (1994)
105. T Aoyama, Y Suzuki, H Ichikawa *J. Med. Chem.* **33** 2583 (1990)
106. B Lučić, N Trinajstić *J. Chem. Inf. Comput. Sci.* **39** 121 (1999)
107. B Lučić, D Amić, N Trinajstić *J. Chem. Inf. Comput. Sci.* **40** 403 (2000)
108. C J Barden, M K Boysworth, F A Palocsay *J. Chem. Inf. Comput. Sci.* **38** 483 (1998)
109. J Meiler, R Meusinger, M Will *J. Chem. Inf. Comput. Sci.* **40** 1169 (2000)
110. B E Mitchell, P C Jurs *J. Chem. Inf. Comput. Sci.* **36** 58 (1996)
111. D L Clouser, P C Jurs *J. Chem. Inf. Comput. Sci.* **36** 168 (1996)
112. O Ivanciuc, J-P Rabine, D Cabrol-Bass, A Panaye, J P Doucet *J. Chem. Inf. Comput. Sci.* **36** 644 (1996)
113. J Grunenberg, R Herges *J. Chem. Inf. Comput. Sci.*, **35** 905 (1995)
114. Q X Guo, S H Luo, H Wang, M S Zhang, Y C Liu *J. Chem. Res. (S)* 38 (1996)
115. L Liu, Q-X Guo *J. Chem. Inf. Comput. Sci.* **39** 133 (1999)
116. M E Brewster, M-J Huang, A Harget, N Bodor *Tetrahedron* **48** 3463 (1992)
117. S Ventura, M Silva, D Pérez-Bendito, C Hervvs *J. Chem. Inf. Comput. Sci.* **37** 287 (1997)
118. A T Balaban, S C Basak, T Colburn, G D Grunwald *J. Chem. Inf. Comput. Sci.* **34** 1118 (1994)
119. L M Egolf, P C Jurs *J. Chem. Inf. Comput. Sci.* **33** 616 (1993)
120. L H Hall, C T Story *J. Chem. Inf. Comput. Sci.* **36** 1004 (1996)
121. A A Gakh, E G Gakh, B G Sumpter, D W Noid, L D Trowbridge *J. Fluorine Chem.* **73** 107 (1995)
122. S Liu, R Zhang, M Liu, Z Hu *J. Chem. Inf. Comput. Sci.* **37** 1146 (1997)
123. E S Goll, P C Jurs *J. Chem. Inf. Comput. Sci.* **39** 974 (1999)
124. I I Baskin, V A Palyulin, N S Zefirov *Dokl. Akad. Nauk* **332** 713 (1993) [a]
125. B E Mitchel, P C Jurs *J. Chem. Inf. Comput. Sci.* **37** 538 (1997)
126. E S Goll, P C Jurs *J. Chem. Inf. Comput. Sci.* **39** 1081 (1999)
127. T Suzuki, R-U Ebert, G Schüürmann *J. Chem. Inf. Comput. Sci.* **37** 1122 (1997)
128. N Bodor, A Harget, M-J Huang *J. Am. Chem. Soc.* **113** 9480 (1991)
129. B E Mitchell, P C Jurs *J. Chem. Inf. Comput. Sci.* **38** 489 (1998)
130. J Huuskonen, M Salo, J Taskinen *J. Chem. Inf. Comput. Sci.* **38** 450 (1998)
131. J M Sutter, P C Jurs *J. Chem. Inf. Comput. Sci.* **36** 100 (1996)
132. H L Engelhardt, P C Jurs *J. Chem. Inf. Comput. Sci.* **37** 478 (1997)
133. B E Mitchell, P C Jurs *J. Chem. Inf. Comput. Sci.* **38** 200 (1998)
134. J M Cense, B Diawara, J J Legendre, G Roullet, in *Proceedings of AIP Conference 330 (E.C.C.C. 1 Computational Chemistry) (F.E.C.S. Conference), Nancy, 1995* p. 556
135. J J Huuskonen, D J Livingstone, I V Tetko *J. Chem. Inf. Comput. Sci.* **40** 947 (2000)
136. A F Duprat, T Huynh, G Dreyfus *J. Chem. Inf. Comput. Sci.* **38** 586 (1998)
137. J Devillers, D Domine, C Guillon, W Karcher *J. Pharm. Sci.* **87** 1086 (1998)

138. H Kränz, V Vill, B Meyer *J. Chem. Inf. Comput. Sci.* **36** 1173 (1996)
139. R Schröder, H Krinz, V Vill, B Meyer *J. Chem. Soc., Perkin Trans. 2* 1685 (1996)
140. H Nefati, J-J Legendre, C Michot, in *Proceedings of AIP Conference 330 (E.C.C.C. 1 Computational Chemistry) (F.E.C.S. Conference), Nancy, 1995* p. 562
141. I I Baskin, A O Ait, N M Halberstam, V A Palyulin, M V Alfimov, N S Zefirov *Dokl. Akad. Nauk* **357** 57 (1997) [a]
142. M Chastrette, D Cretin, C El Aidi *J. Chem. Inf. Comput. Sci.* **36** 108 (1996)
143. O Ivanciuc *Rev. Roum. Chim.* **40** 567 (1995)
144. V Kvasnicka, S Sklenak, J Pospichal *Theor. Chim. Acta* **86** 257 (1993)
145. N Ghoshal, S N Mukhopadhayay, T K Ghoshal, B Achari *Indian J. Chem., Sect. B* **32** 1045 (1993)
146. S-S So, M Karplus *J. Med. Chem.* **39** 5246 (1996)
147. M Jalali-Heravi, F Parastar *J. Chem. Inf. Comput. Sci.* **40** 147 (2000)
148. M Wiese *Quantum Struct.-Act. Relat.* **10** 369 (1991)
149. T Aoyama, Y Suzuki, H Ichikawa *J. Med. Chem.* **33** 905 (1990)
150. J N Weinstein, K W Kohn, M R Grever, V K Viswanadhan, L V Rubinstein, A P Monks, D A Scudiero, L Welch, A D Koutsokos, A J Chiausa, K D Paull *Science* **258** 447 (1992)
151. S Hatrik, P Zahradnik *J. Chem. Inf. Comput. Sci.* **36** 992 (1996)
152. R Vendrame, R S Braga, Y Takahata, D S Galvao *J. Chem. Inf. Comput. Sci.* **39** 1094 (1999)
153. G Gini, M Lorenzini *J. Chem. Inf. Comput. Sci.* **39** 1076 (1999)
154. N Ghoshal, S N Mukhopadhayay, T K Ghoshal, B Achari *Bioorg. Med. Chem. Lett.* **3** 329 (1993)
155. M Brinn, P T Walsh, M P Payne, B Bott *SAR QSAR Environ. Res.* **1** 169 (1993)
156. I K Lyubimova, S K Abilev, N M Halberstam, I I Baskin, V A Palyulin, N S Zefirov *Izv. Akad. Nauk, Ser. Biol.* 180 (2001)
157. G W Kauffman, P C Jurs *J. Chem. Inf. Comput. Sci.* **40** 753 (2000)
158. M D Wessel, P C Jurs, J W Tolan, S M Muskal *J. Chem. Inf. Comput. Sci.* **38** 726 (1998)
159. J Jaén-Oltra, T Salabert-Salvador, F J Garcia-March, F Pérez-Giménez, F Tomás-Vert *J. Med. Chem.* **43** 1143 (2000)

[a] — *Dokl. Chem. (Engl. Transl.)*