Predicting Carcinogenicity of Polycyclic Aromatic Hydrocarbons from Back-Propagation Neural Network

Didier Villemin,* Driss Cherqaoui, and Abdelhalim Mesbah

Ecole Nationale Supérieure d'Ingénieurs de Caen (E.N.S.I. de Caen), I.S.M.R.A., U.R.A. 480 CNRS, 6 boulevard du Maréchal Juin, 14050 Caen Cedex, France

Received July 11, 1994®

Models of relationships between structure and carcinogenicity of polycyclic aromatic hydrocarbons were constructed by means of a multilayer neural network using the back-propagation algorithm. The molecular descriptors used were derived from graph theory. The neural network (NN) was used to classify the compounds studied into two categories, namely inactive or active. To evaluate the predictive power of an NN model, the cross-validation procedure was used. The total prediction accuracy of 86% (90% of the actives correctly identified) provided an evidence of the usefulness of the present neural algorithm.

INTRODUCTION

Polycyclic aromatic hydrocarbons (PAHs) have become ubiquitous in the environment.[1] A large variety of combustion processes from waste incineration,[2] coal[3] wood, biomass burning,[4] and traffic exhaust[5] spray PAHs in the environment. At the turn of the century, high incidence of cancer among chimney sweeps was related to PAHs. One of the earliest attempts to relate structures and carcinogenicity was proposed by B. and A. Pullman[6] in 1955. From Hückel's calculations they related carcinogenicity of PAHs to the existence of an active K region and an inactive L region in PAHs (Figure 1).

Several biochemists showed that in many cases the metabolism of PAHs proceeded via different routes involving a benzene ring. A typical example is benzo[*a*]pyrene: after initial oxidation at 7,8-position by cytochrome P450, the molecule is transformed into diol epoxide (Figure 2). This epoxide is believed to react with a nucleic base and give rise to an event conducive to cancer.[7−9] As can be seen from this mechanism, the propensity of parent PAHs to be carcinogenic depends on the ability to form the cation that arylates the nucleic base. Jerina et al.[10] found that stabilization of the "bay region" (Figure 1) cation is directly proportional to the carcinogenicity of PAHs. In this approach the carcinogenicity can be related to several theoretical reactivity indices[11,12] or physical quantity such as the ionization potential[13] or the averaged ^{13}C chemical shifts over all the carbons in the ring structures of aromatic compounds.[14]

Structure−activity relationships of PAHs have been studied using both empirical and theoretical methods. Thus Hansch and Fugita[15] correlated the logarithm of the carcinogenicity of 47 PAHs with a quadratic term in the logarithm of the 1-octanol/water partition coefficient and with the total charge of the K region. Using the graph theory, Lall[16] has showed that the carcinogenic activity of PAHs depends upon the symmetry and the relative position of K-bonds in the molecules. Nordén et al.[17] have studied the relationship between the carcinogenic activity and the structure of 32 unsubstituted PAHs by the SIMCA pattern recognition.

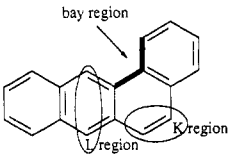


Figure 1. L and K region in benzo[*a*]anthracene.

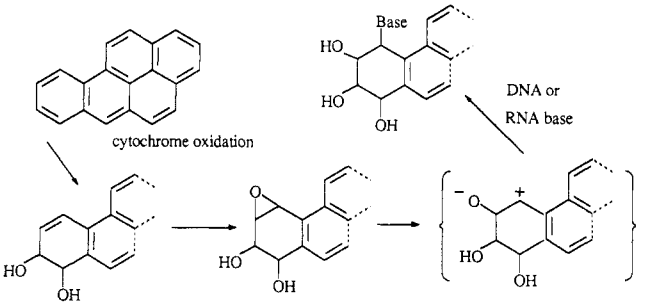


Figure 2. Metabolism of benzo[*a*]pyrene.

Yuan et al. have used a combination of parameters which pertain to molecular size and shape, molecular connectivity indices, and indicator variables to model the carcinogenic activity of many compounds including PAHs.[18]

NEURAL NETWORK

Artificial NNs or NNs for short are mathematical models of biological neural systems. Three components constitute an NN: the processing elements, the topology of the connections between the nodes, and the learning rule. In this paper, the specific algorithm used is the back-propagation system. Its goal is to minimize an error function. A description of the back-propagation algorithm was given previously,[19] and a more extensive description can be found in another work.[20,21]

NNs have recently[22,23] become the focus of much attention largely because of their wide range of applicability and the ease with which they can handle complex and nonlinear problems. A leading reference book[24] on the application and the meaning of NNs in chemistry has been published by

* Author to whom all correspondence should be addressed.
® Abstract published in *Advance ACS Abstracts,* August 15, 1994.

0095-2338/94/1634-1288$04.50/0

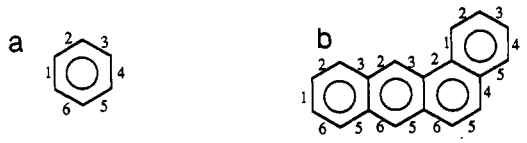


Figure 3. (a) A boundary code for benzene is 123456. The labels 1−6 indicate edges of the benzene graph. (b) An illustrative example for construction of a BC of benzo[*a*]anthracene.

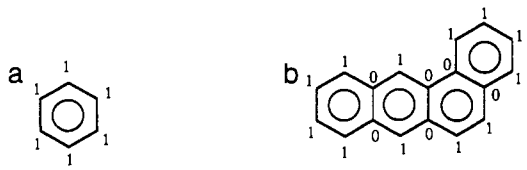


Figure 4. (a) The PC-1 for benzene is 111111. (b) An illustrative example for construction of the PC-1 of benzo[*a*]anthracene.

Zupan and Gasteiger in which an extensive list of references can be found. NNs have been applied to the identification of proton-NMR spectra,[25] the interpretation of IR spectra,[26,27] the prediction of ^{13}C chemical shifts,[28] the classification of mass spectra,[29] the estimation of aqueous solubilities,[30] the determination of protein structure,[31,32] the investigation of quantitative structure−Activity relationships (QSAR),[33−36] and the prediction of chemical reactivity.[37,38]

GRAPH THEORY

The carbon skeletons of the polycyclic benzenoid hydrocarbon will be represented by graphs, called benzenoid graphs or polyhexes[39] for short. In this study we consider only planar polyhexes without holes,[40] and their structures can be represented by their boundaries, their interiors being easily deduced.

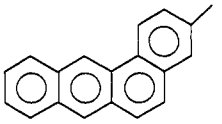
Boundary Code. The boundary code (BC)[41] is a numerical representation of the polyhex boundary. For a single hexagon (benzene) one can denote each edge by a digit (1−6). The digit 1 is arbitrarily assigned to the left vertical edge of the hexagon (Figure 3−a). Similarly the walk in the boundary is to be clockwise. Thus a BC of benzene is "123456". In this case all bonds are on the boundary. However for any other polyhex, several bonds belong to the interior of a polyhex and are not listed in the BC.

For example, consider the polyhex in Figure 3b and start by the most left edge; the sequence obtained is S1: "123223123454565656". Of course this sequence of digits is not unique, but by cyclic permutations of S1 and those of the inverted sequence S2: "656565454321232321", several numerical representations can be given. If we apply the same process used for S1 and S2 to the sequences obtained by changing in Figure 3b the numbering of the edges so that digit *1* is assigned once and only once to each edge, all numerical representations of this polyhex can be generated. We chose the BC which is the lexicographic maximum[42] of all sequences of a fixed polyhex.

Perimeter Code. For planar polyhex C. W. Herndon et al.[43a,b] have assigned to each carbon atom the values "1" or "0" according to its having one or zero hydrogen atom; then a single hexagon (Figure 4a) is represented by the sequence 111111. But in all other cases, several numerical representations will arise for each planar polyhex:benzo[*a*]anthracene (Figure 4b) described by several sequences, depending upon the vertex at which the code starts. Among all the sequences, the one chosen is the lexicographic maximum, called PC-1.

Table 1. Examples of BC, PC-1, and PC-2 for Three Molecules

molecule		codes
benzo[*a*]anthracene	BC:	656561612343232345
	PC-1:	111101101011110100
	PC-2:	421410
benzo[*c*]phenanthrene	BC:	656161234321234545
	PC-1:	111101101101111000
	PC-2:	422400
benzo[*g*]chrysene	BC:	6561656123212345432345
	PC-1:	111101100111001111000
	PC-2:	42040400



Construction of MPC-2
First step : Find the PC-2 without Me
PC-2 : 4 2 1 4 1 0
Second step : Increase all digits by 4, except 0
(8 6 5 8 5 0)
Second step : Give the Me digit '1', '2', '3' or '4' acording to its position in the cluster.
MPC-2 : 8 3 6 5 8 5 0

Construction of MBC
First step : Find the BC without Me
BC : 6565616123432323345
Second step : Give the Me digit '7'. Insert this digit between those corresponding to the two bonds on either sides of this Me.
MBC : 65656161723432323345

Figure 5. Steps for construction of the MBC and MPC-2 of 3-methylbenz[*a*]anthracene.

For polyhexes with more than one benzenoid ring and from PC-1, the authors obtained a much shorter code (PC-2) by replacing sequences of digits in the PC-1 as follows: '11110' by 4, '1110' by 3, '110' by 2, and '10' by 1. Examples of BC, PC-1, and PC-2 for three molecules are listed in Table 1. It is clear that the BC and the PC can code only nonsubstituted PAHs, and thus cannot be applied to all the compounds studied (Table 2). In order to avoid such a restriction we have used slightly modified BC and PC. Figure 5 shows this modification. Examples of modified BC (MBC) and PC-2 (MPC-2) for three molecules are listed in Table 3. For the modified BC with less than 26 numbers or for the modified PC-2 with less than 10 numbers, the rest of the numeric code was filled out with zeros to ensure that all the input patterns had the same number of elements. In other words, all modified BC and PC-2 had 26 and 10 numbers, respectively.

METHOD

The compounds (Table 2) were taken from articles by Dipple,[44,45] Cavalieri et al.,[13] and Richard et al.[46] In the Dipple data set, many PAHs are arranged according to their carcinogenic potency into four classes: inactive, slight, moderate, and high. Since a certain number of compounds and their activities were not present in Dipple's article, we took them from Cavalieri and Richard's articles.

In some cases, compounds were classified into several categories because different results were obtained with different species or with different methods. For example, 5-methylbenzo[*a*]anthracene is classified as an inactive compound when it is injected intramuscularly into rats. It is classified as a slight carcinogen when it is either painted on mouse skin or subcutaneously injected into rats. Finally it is classified as a moderate carciongen when it is subcutaneously injected into mice.

Because it is not possible to classify correctly the compounds used in this study in different categories, we

Table 2. Structures and Activities of PAHs (+ for Carcinogenic Compounds and − for Noncarcinogenic Ones), Experimental Activities, and Predicted Activities by Neural Network

	name of polycyclic aromatic hydrocarbon	exp activities	NN predicted	ref		name of polycyclic aromatic hydrocarbon	exp activities	NN predicted	ref
1	naphthalene	−	−	13, 45	48	11-methylbenz[*a*]anthracene	+	+	45, 46
2	anthracene	−	−	13, 45	49	12-methylbenz[*a*]anthracene	+	+	13, 45, 46
3	phenanthrene	−	−	13, 45, 46	50	1,12-dimethylbenz[*a*]anthracene	−	−	45, 46
4	naphthacene	−	+	13, 45, 46	51	7,12-dimethylbenz[*a*]anthracene	+	+	13, 45, 46
5	benzo[*a*]anthracene	+	+	13, 45, 46	52	6,12-dimethylbenz[*a*]anthracene	+	+	13, 45
6	chrysene	+	+	13, 45, 46	53	6,8-dimethylbenz[*a*]anthracene	+	+	13, 45
7	benzo[*c*]phenanthrene	+	+	13, 45, 46	54	6,8,12-trimethylbenz[*a*]anthracene	+	−	13, 45
8	pyrene	−	−	13, 45, 46	55	1-methylchrysene	+	+	45, 46
9	triphenylene	−	−	13, 45, 46	56	4-methylchrysene	+	+	45, 46
10	dibenz[*a,h*]anthracene	+	+	13, 45, 46	57	5-methylchrysene	+	+	13, 45, 46
11	dibenz[*a,j*]anthracene	+	+	13, 45, 46	58	6-methylchrysene	+	+	45
12	benzo[*a*]pyrene	+	+	13, 45, 46	59	1-methylbenzo[*c*]phenanthene	+	+	13
13	benzo[*c*]chrysene	+	+	45, 46	60	2-methylbenzo[*c*]phenanthrene	+	+	45, 46
14	benzo[*g*]chrysene	+	+	45, 46	61	3-methylbenzo[*c*]phenanthrene	+	+	13, 46
15	benzo[*e*]pyrene	+	+	46	62	4-methylbenzo[*c*]phenanthrene	+	+	13, 46
16	benzo[*a*]naphthacene	−	−	45, 46	63	5-methylbenzo[*c*]phenanthrene	+	+	13, 45, 46
17	pentaphene	−	−	45, 46	64	6-methylbenzo[*c*]phenanthrene	+	+	13, 46
18	dibenz[*a,c*]anthracene	+	+	13	65	5,8-dimethylbenzo[*c*]phenanthrene	−	+	46
19	pentacene	−	−	45, 46	66	1-methylpyrene	−	−	13, 46
20	dibenzo[*c,g*]phenanthrene	−	+	45, 46	67	2-methylpyrene	−	−	13, 45, 46
21	picene	−	+	13, 45, 46	68	4-methylpyrene	−	−	13, 46
22	benzo[*b*]chrysene	−	−	45	69	1-methylbenzo[*a*]pyrene	+	+	13
23	dibenzo[*b,g*]phenanthrene	−	−	45, 46	70	2-methylbenzo[*a*]pyrene	+	+	13, 45
24	perylene	−	−	45, 46	71	3-methylbenzo[*a*]pyrene	+	+	13, 45
25	anthanthrene	−	−	45	72	4-methylbenzo[*a*]pyrene	+	+	13, 45
26	naphtho[1,2-*b*]triphenylene	−	−	45, 46	73	5-methylbenzo[*a*]pyrene	+	+	13, 45
27	dibenzo[*e,l*]pyrene	−	−	13, 45	74	6-methylbenzo[*a*]pyrene	+	+	13, 45
28	dibenzo[*a,e*]pyrene	+	+	45	75	7-methylbenzo[*a*]pyrene	+	+	13, 45
29	anthra[1,2-*a*]anthracene	−	−	45, 46	76	8-methylbenzo[*a*]pyrene	−	+	45
30	dibenzo[*b,k*]chrysene	−	−	45, 46	77	10-methylbenzo[*a*]pyrene	+	−	13
31	naphtho(2,3-*a*)pyrene	+	−	45, 46	78	11-methylbenzo[*a*]pyrene	+	+	13, 45
32	benzo[*c*]penthaphene	−	−	45, 46	79	12-methylbenzo[*a*]pyrene	+	+	13, 45
33	dibenzo[*a,l*]naphthacene	−	−	17	80	7,10-dimethylbenzo[*a*]pyrene	−	+	13
34	dibenzo[*a,j*]naphthacene	−	−	45, 46	81	4,5-dimethylbenzo[*a*]pyrene	+	+	13, 45
35	dibenzo[*a,l*]pyrene	+	+	13, 45	82	1,2-dimethylbenzo[*a*]pyrene	+	+	13, 45
36	dibenzo[*a,h*]pyrene	+	+	13, 45, 46	83	2,3-dimethylbenzo[*a*]pyrene	+	+	13, 45
37	dibenzo[*a,i*]pyrene	+	+	13, 45	84	3,12-dimethylbenzo[*a*]pyrene	+	+	13, 45
38	1-methylbenz[*a*]anthracene	−	−	13, 45, 46	85	1,3-dimethylbenzo[*a*]pyrene	+	+	13, 45
39	2-methylbenz[*a*]anthracene	−	−	13, 45, 46	86	1,4-dimethylbenzo[*a*]pyrene	+	−	13, 45
40	3-methylbenz[*a*]anthracene	−	−	13, 45, 46	87	1,6-dimethylbenzo[*a*]pyrene	+	+	13, 45
41	4-methylbenz[*a*]anthracene	−	−	45, 46	88	3,6-dimethylbenzo[*a*]pyrene	+	+	13, 45
42	5-methylbenz[*a*]anthracene	−	+	13, 45, 46	89	6-methylanthanthrene	+	+	13
43	6-methylbenz[*a*]anthracene	+	+	13, 45, 46	90	6,12-dimethylanthanthrene	+	+	13
44	7-methylbenz[*a*]anthracene	+	+	13, 45, 46	91	9,10-dimethylanthracene	+	+	13, 46
45	8-methylbenz[*a*]anthracene	+	+	45, 46	92	tribenzo[*a,e,i*]pyrene	+	−	13, 46
46	9-methylbenz[*a*]anthracene	+	+	45, 46	93	coronene	−	−	13, 46
47	10-methylbenz[*a*]anthracene	+	+	45, 46	94	benzo[*g,h,i*]perylene	+	−	45, 46

Table 3. Examples of Modified Boundary Code (MBC) and Modified Perimeter Code (MPC-2) for Three Molecules

molecule	codes	
benzo[*a*]anthracene	MBC:	656561612343232345
	MPC-2:	865850
3-methylbenz[*a*]anthracene	MBC:	6565616172343232345
	MPC-2:	8365850
6,8,12-trimethylbenz[*a*]-anthracene	MBC:	656576161234327323457
	MPC-2:	862581510

reduced the number of classes defined by Dipple from 4 to 2: carcinogenic and noncarcinogenic. A compound was considered to be noncarcinogenic if any only if it was classified as inactive; otherwise it was classified as carcinogenic. Thus, among all the compounds studied, 59 were carcinogenic and 35 were noncarcinogenic.

Each molecule is coded by two input patterns corresponding to the modified BC and PC-2. The output of NNs which describes the activity is coded 1 or (1,0) if the molecule is carcinogenic and 0 or (0,1) otherwise.

We used a network with 10 or 26 units and a bias in the input layer, a variable hidden layer including bias, and one or two units in the output layer. Input and output data were normalized between 0.1 and 0.9. The weights were initialized to random values between −0.5 and +0.5, and no momentum was added. The learning rate was initially set to 1 and was gradually decreased until the error function could no longer be minimized. The maximum number of iterations was set to 2000, but sufficient convergence was usually obtained, at least for small networks, after 800 iterations.

All computations were performed on an Iris Indigo (Silicon Graphics) workstation using our own programs, written in C language.

RESULTS AND DISCUSSION

In a back-propagation NN the input and output neurons are known since they represent, respectively—in this study—the modified PC-2 (or the modified BC) and the activity of the

Table 4. Comparison of % of Correct Classification with Different ϱ Values[a]

NN's configuration	ϱ	% of correct classification
26-1-1	3.24	88
26-2-1	1.65	89
26-3-1	1.11	97
26-4-1	0.83	98

[a] The modified BC was used as input pattern. Four networks were constructed with configuration 26-x-1, where $x = 1-4$ (x, number of hidden neurons).

Table 5. Comparison of % of Correct Classification with Different ϱ Values[a]

NN's configuration	ϱ	% of correct classification
10-1-1	7.23	90
10-2-1	3.76	96
10-3-1	2.54	98
10-4-1	1.92	98
10-5-1	1.54	99
10-6-1	1.29	99

[a] The modified PC-2 was used as input pattern. Six networks were constructed with configuration 10-x-1, where $x = 1-6$.

molecules. Unfortunately, there are neither theoretical results available nor satisfying empirical rules that would enable us to determine the number of hidden layers and of neurons contained in these layers. However, for most of the applications of NNs to chemistry, one hidden layer seems to be sufficient. For the determination of the number of hidden neurons, we have recently discussed[19] the usefulness of a ϱ parameter defined as

$$\varrho = \frac{\text{number of data point in the training set}}{\text{sum of the number of connections in the NN}}$$

According to Zupan and Gasteiger[47] "a good rule of thumb is that the number of data values taken for training should be equal to or greater than the number of weights to be determined in the network" (i.e., $\varrho \geq 1$). In this paper, different architectures of NN have been tried, and two studies have been achieved: classification and prediction. The term classification is used when the NN estimates the activity for molecules in the training set. When it estimates the activity for molecules not included in the training set, this is prediction.

Classification Ability of the NNs. The three-layer NN has been adapted by standard back-propagation method for all 94 compounds (Table 2). After the training phase (2000 training cycles), the classification ability of the NN was tested on the same 94 compounds. This training process was performed many times with different architectures of NNs. Since the method used for representing the chemical information for the NN is a major factor in determining the predictive ability of the network, two different approaches were used in this study. The first approach, the most successful, employed information derived from the modified PC-2 of PAHs. The second one employed the modified BC. Only one output is used to code the carcinogenic activity.

The results of this analysis are given in Tables 4 and 5. Observe that the classification performance of the NNs generally increases at the same time as the number of hidden units since the use of more hidden units provides a greater flexibility in the network encoding process.

Table 6. Comparison of % of Correct Prediction with Different ϱ Values[a]

NN's configuration	ϱ	% of correct classification	% of correct classification for carcinogenic compds
10-2-1	3.72	77	85
10-3-1	2.51	82	88
10-4-1	1.90	76	80
10-5-1	1.52	74	78
10-6-1	1.27	71	73

[a] The modified PC-2 was used as input pattern. Only one output neuron is used. Five networks were constructed with configuration 10-x-1, where $x = 2-6$.

Table 7. Comparison of % of Correct Prediction with Different ϱ Values[a]

NN's configuration	ϱ	% of correct classification	% of correct classification for carcinogenic compds
10-2-2	3.32	81	86
10-3-2	2.27	85	90
10-4-2	1.72	86	90
10-5-2	1.39	77	78
10-6-2	1.16	74	76

[a] The modified PC-2 was used as input pattern. Two output neurons were used. Five networks were constructed with configuration 10-x-2, where $x = 2-6$.

Results comparing the modified PC-2 and BC indicate that the former representation provides better classification than the latter. Thus for the NN prediction ability, only the modified PC-2 is chosen. Another reason to choose only PC-2 is that the NN using the modified PC-2 as input pattern has fewer weights to adjust than that using the BC, with the same number of hidden neurons.

Prediction Ability of the NNs. One of the most important attributes of NNs is their ability to generalize, that is their ability to make reliable predictions on new data with similar accuracy to that obtained with the training set. Having determined the range of hidden units giving good classification of the data set, we turned to the more important predictive aspect of QSAR. To determine the predictive ability of the selected network, cross-validation has been used. In this process one compound was removed from the data set, and the remaining 93 compounds served as the training set for the network. After the training, the parameters of the compound unknown by the network were put into the network, and the predicted activity of this compound was evaluated. This procedure was repeated 94 times, and the predicted activities of the entire data set were obtained.

We first approached the problem of classifying compounds as carcinogenic or noncarcinogenic by using only one output. This was coded as 0 for noncarcinogenic compounds or 1 for carcinogenic ones. Our second task was to use two outputs instead of one, one for each class. For the carcinogenic compounds the output pattern is (1,0) and (0,1) for the noncarcinogenic ones.

The results of this analysis are shown in Tables 6 and 7. These results are satisfying and show that with a simple coding system, the NNs give correct previsions. The NNs were able to extract information from examples to develop an internal representation of the carcinogenic activity of PAHs without explicity incorporating rules into the network. We also divided randomly the whole set of patterns into two

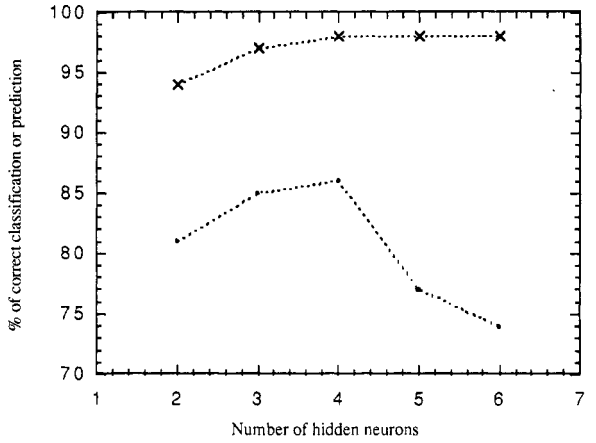


Figure 6. Classification ability of NN (top curve) and predictive ability of NN (bottom curve).

disjoint subsets, called the training set (64 molecules) and the test set (30 molecules). Out of the 30 molecules of the test set, only five (16.7%) were wrongly predicted by the NN with the configuration 10-3-2.

With two units in the output layer, the predictive ability of the network is more important than with only one output. This can be explained by the fact that if we pass from one to two output neurons, we increase the number of connections between the hidden layer and the output layer. That increase has enabled us to get an NN with sufficient connections to stock the information it needs in order to separate carcinogenic compounds from noncarcinogenic ones. However, when too many connections are used, for example when the number of hidden neurons is increased, the prediction ability decreases (Figure 6). The behavior observed in Figure 6 is expected because when the number of weights increases and reaches a certain threshold depending upon the number of different samples in the training set, overtraining appears. The overtraining effect can be explained as a consequence of parameter redundancy, that is the NN has more parameters than are needed for the resolution of the problem.

Among all architectures of NNs, the best one is 10-4-1. The results of QSAR done by this NN are listed in Table 2. Out of the 94 molecules of the test set only 13 (14%) were wrongly predicted. Among these 13 compounds, six were predicted as being noncarcinogenic even if they were carcinogenic. In this study we have shown that in the NN applications, it is not necessary to use another type of descriptors based on measured variables to describe the molecular structure. NNs are endowed with a very valuable feature because they are able to extract the basic information directly from boundaries of PAHs. It is not always necessary to use parameters that have been created by sophisticated models as descriptors for NN applications. The molecular structure should be determined in a simple way.

CONCLUDING REMARKS

This paper has discussed the use of back-propagation NNs for predicting the carcinogenic activity of PAHs. After the NN had been fully trained, it was shown that it was capable of forming reliable generalizations to classify PAHs (according to their activities) than it had never seen. In other words, a properly trained NN can be used to rapidly classify PAHs with a high accuracy rate.

A simple modified PC-2 provided enough information for predicting this activity. Using these boundary descriptors, the NNs gave good prediction accuracy. Because of its ability to identify nonlinear relationships, back-propagation is a very promising neural algorithm for the determination of quantitative structure−activity relationships and as such it is a valuable tool for the medicinal chemist.

REFERENCES AND NOTES

(1) Grimmer, G. Ed., *Environmental Carcinogens, Polycyclic Aromatic Hydrocarbons, Chemistry, Occurrence, Biochemistry, Carcinogenicity*; CRC Press Inc.: Boca Raton, FL, 1983.
(2) Taylor, P.; Dellinger, B.; Lee, C. C. Development of a thermal stability-based ranking of hazardous organic compound incinerability. *Environ. Sci. Technol.* **1990**, *24*, 316−328.
(3) Grimmer, G.; Naujac, K. W.; Dettbarn, G.; Brune, H.; Deutsch-Wenzel, R.; Misfeld, J. Analysis of balance of carcinogenic impact from emission condensates of automobile exhaust, coal heating, and used engine oil by mouse skin painting as a carcinogen specific detector. *Polynucleic Aromatic Hydrocarbons: Physics, Biology, and Chemistry 6th International Symposium*; Cooke, M., Dennis, A. J., Fisher, G. L., Eds.; Bartelle Press: Columbus, OH, 1981; p 335.
(4) Freeman, D. J.; Cattell, F. C. R. Woodburning as a source of atmospheric polycyclic aromatic hydrocarbons. *Environ. Sci. Technol.* **1990**, *24*, 1581−1585.
(5) Takada, H.; Onda, T.; Ogura, N. Determination of polyclic aromatic hydrocarbons in urban street dusts and their source materials by capillary gas chromatography. *Environ. Sci. Technol.* **1990**, *24*, 1179−1186.
(6) Pullman, A.; Pullman, B. Electronic structure and carcinogenic activity of aromatic molecules. *Ad. Cancer Res.* **1955**, *3*, 117−169.
(7) Ulrich, V.; Roots, I.; Hildebrant, A. G.; Eatabrook, R. K. W. *Microsomes and Drug Oxidations*; Conney, A. H., Ed.; Pergamon Press, Inc.: Oxford, 1977.
(8) Geacintov, N. E. Reaction pathways of benzo[*a*]pyrene diol epoxide and DNA. *Polynuclear Aromatic Hydrocarbons: Physics, Biology, and Chemistry 6th International Symposium*; Cooke, M., Dennis, A. J., Fisher, G. L., Eds.; Bartelle Press: Columbus, OH, 1981; p 311.
(9) Hodgson, R. M.; Cary, P. D.; Grover, P. L.; Sims, P. Metabolic activation of chrysene by hamster embryo cells: evidence for the formation of a "bay-region" diol-epoxide-N^2-guanine adduct in RNA. *Carcinogenesis (London)* **1983**, *4*, 1153.
(10) Jerina, D. M.; Lehr, R. E. in ref 7, p 709.
(11) Berger, D.; Smith, A.; Seybold, G.; Serve, M. P. Correlation of electronic reactivity index with carcinogenicity in Polycyclic Aromatic Hydrocarbons. *Tetrahedron Lett.* **1978**, *3*, 231−234.
(12) Sjentplay, L. V. Carcinogenesis by Polycyclic Aromatic Hydrocarbons: A Multilinear Regression on New Type PMO Indices. *J. Am. Chem. Soc.* **1984**, *106*, 6021−6028.
(13) Cavalieri, E. L.; Rogan, E. G.; Roth, R. W.; Saugier, R. K.; Hakam, A. The Relationship Between Ionization Potential and Horseradish Peroxidase/Hydrogen Peroxide−Catalysed Binding of Aromatic Hydrocarbons to DNA. *Chem. Biol. Interactions* **1983**, *47*, 87−109.
(14) Sakamoto, Y.; Watanabe, S. On the relationship between the chemical structure and the carcinogenicity of Polycyclic and Chlorinated Monocyclic Aromatic Compounds as studied by means of ^{13}C NMR. *Bull. Chem. Soc. Jpn.* **1986**, *59*, 3033−3038.
(15) Hansch, C.; Fujita, T. A Method for the Correlation of Biological Activity and Chemical Structure. *J. Am. Chem. Soc.* **1964**, *86*, 1616−1626.
(16) Lall, R. S.; Structure-Activity Relationship in Hydrocarbon Carcinogens. *Match.* **1984**, *15*, 251−261.
(17) Nordén, B.; Edlund, U.; Wold, S. Carcinogenicity of Polycyclic Aromatic Hydrocarbons Studied by SIMICA Pattern Recognition. *Acta Chem. Scand.* **1978**, *B32*, 602−608.
(18) Klopman, G. Artificial Intelligence Approach to Structure-Activity Studies. Computer Automated Structure Evaluation of Biological Activity of Organic Molecules. *J. Am. Chem. Soc.* **1984**, *106*, 7315−7321.
(19) Cherqaoui, D.; Villemin, D. Use of a Neural Network to determine Boiling Point of Alkanes. *J. Chem. Soc., Faraday Trans.* **1994**, *90*, 97−102.
(20) McClelland, J. L.; Rumelhart, D. E. *Parallel Distributed Processing*; McClelland, J. L., Rumelhart, D. E., and the PDP Research Group, Eds., MIT Press: Cambridge, MA, 1988; Vol. 1, pp 319−362.
(21) Freeman, J. A.; Skapura, D. M. *Neural Networks Algorithms, Applications, and Programming Techniques*; Addition-Wesley Publishing Company: Reading, PA, 1991; pp 89−125.
(22) Burns, J. A.; Whitesides, G. M. Feed-Forward Neural networks in Chemistry: Mathematical Systems for Classification and Pattern Recognition. *Chem. Rev.* **1993**, *93*, 2583−2601.
(23) Tusar, M.; Zupan, J.; Gasteiger, J. Neural networks and modelling in chemistry. *J. Chim. Phys.* **1992**, *89*, 1517−1529.

(24) Zupan, J.; Gasteiger, J. In *Neural Networks for chemists*; VCH Publishers: New York, 1993.
(25) Thomsen, J. U.; Meyer, B. Pattern Recognition of the ^{1}HNMR Spectra of Sugar Alditols Using a Neural Network. *J. Magn. Reson.* **1989**, *84*, 212−217.
(26) Robb, E. W.; Munk, M. E. A Neural Network Approach to Infrared Spectrum Interpretation. *Mikrochim Acta [Wien]* **1990**, *I*, 131−155.
(27) Munk, M. E.; Madison, M. S.; Robb, E. W. Neural Network Models to Infrared Spectrum Interpretation. *Mikrochim Acta [Wien]* **1991**, *II*, 505−514.
(28) Kvasnicka, V. An Application of Neural Networks in Chemistry. Prediction of ^{13}C NMR Chemical Shifts. *J. Math. Chem.* **1991**, *6*, 63−76.
(29) Curry, B.; Rumelhart, D. E. MS net: A Neural Network That Classifies Mass Spectra *Tetrahedron Comput. Methodol.* **1990**, *3*, 213−237.
(30) Bodor, N.; Harget, A.; Huang, M. J. Neural Network Studies. 1. Estimation of the Aqueous Solubility of Organic Compounds. *J. Am. Chem. Soc.* **1991**, *113*, 9480−9480.
(31) Holley, L. H.; Karplus, M. Protein Secondary Structure Prediction with a Neural Network. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 152−156.
(32) Qian, N.; Sejnowski, T. J. Predicting the Secondary Structure of Globular Proteins Using Neural Network Models. *J. Mol. Biol.* **1988**, *202*, 865−884.
(33) Villemin, D.; Cherqaoui, D.; Cense, J. M. Neural Networks Studies: Quantitative Structure-Activity Relationship of Mutagenic Aromatic Nitro Compounds. *J. Chim. Phys.* **1993**, *90*, 1505−1519.
(34) Aoyama, T.; Ichikawa, H. Basic Operating of Characteristics of Neural Networks When Applied to Structure-Activity Studies. *Chem. Pharm. Bull.* **1991**, *39*, 358−366.
(35) Adler, B.; Ammon, K.; Dobers, S.; Winterstein, M.; Ziesmer, H. Computer aided prediction of carcinogenic properties of aminobiphenyls. *Chem. Technik.* **1992**, *44*, 363−367.
(36) Domine, D.; Devillers, J.; Chastrette, M.; Karcher, W. Estimating Pesticide Field Half-lives from a Back-Propagation Neural Network. *SAR QSAR Environ. Res.* **1993**, *1*, 211−219.
(37) Simon, V.; Gasteiger, J.; Zupan, J. A Combined Application of Two Different Neural Network Types For the Prediction of Chemical Reactivity. *J. Am. Chem. Soc.* **1993**, *115*, 9148−9159.
(38) Elrod, D. W.; Maggiora, G. M.; Trenary, R. G. Applications of Neural Networks in Chemistry. 1. Prediction of Electrophilic Aromatic Substitution Reactions. *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 477−484.
(39) Balaban, A. T. *Chemical Applications of Graph Theory*; Balaban, A. T., Ed.; Academic: New York, 1976; pp 63−102.
(40) Knop, J. V.; Szymanski, K.; Jericevic, Z.; Trinajstic, N. Computer Enumeration and Generation of Benzenoid Hydrocarbons and Identification of Bay Regions. *J. Comput. Chem.* **1983**, *4*, 23−32.
(41) Trinajstic, N.; Jericevic, Z.; Knop, J. V.; Müller, W. R.; Szymanski, K. Computer Generation of Isomeric Structures. *Pure Appl. Chem.* **1983**, *55(2)*, 379−390.
(42) Knop, J. V.; Müller, W. R.; Jericevic, Z.; Trinajstic, N. Computer Enumeration and Generation of Trees and Rooted Trees. *J. Chem. Inf. Comput. Sci.* **1981**, *21*, 91−99.
(43) (a) Herndon, W. C.; Bruce, A. J. Perimeter Codes for Benzenoid Aromatic Hydrocarbons; In *Graph Theory and Topology in Chemistry*; King, R. B., Rouvray, D. H., Eds.; Elsevier Science: Amsterdam, 1987; Vol. 51, pp 491−563. (b) Herndon, W. C.; Bruce, A. J. Linear Notation for Benzenoid Aromatic Hydrocarbons. Molecular Similarity Based on Notation Similarity. *J. Math. Chem.* **1988**, *2*, 155−169.
(44) Dipple, A. *Polynuclear aromatic carcinogens, Chemical Carcinogens*; Searle, C. E., Ed.; American Chemical Society: Washington, DC, 1976; pp 245−314.
(45) Dipple, A.; Moschel, R. C.; Bigger, C. A. H. *Polynuclear aromatic carcinogens, Chemical Carcinogens*; 2nd ed.; Searle, C. E., Ed.; American Chemical Society: Washington, DC, 1984; vol. 1, pp 41−163.
(46) Richard, A. M.; Woo, Y. A. CASE-SAR analysis of polycyclic aromatic hydrocarbon carcinogenicity. *Mutation Res.* **1990**, *242*, 285−303.
(47) Zupan, J. Gasteiger, J. *Neural Networks for Chemists*; VCH Publishers: New York, 1993, pp 263−265.