# Predictive Carcinogenicity: A Model for Aromatic Compounds, with Nitrogen-Containing Substituents, Based on Molecular Descriptors Using an Artificial Neural Network

Giuseppina Gini* and Marco Lorenzini

Dipartimento di Elettronica e Informazione, Politecnico di Milano, I-20133 Milano, Italy

Emilio Benfenati, Paola Grasso, and Maurizio Bruschi

Department of Environmental Health Sciences, Istituto di Ricerche Farmacologiche "Mario Negri", I-20157 Milano, Italy

Received April 25, 1999

A back-propagation neural network to predict the carcinogenicity of aromatic nitrogen compounds was developed. The inputs were molecular descriptors of different types: electrostatic, topological, quantum-chemical, physicochemical, etc. For the output the index TD50 as introduced by Gold and colleagues was used, giving a continuous numerical parameter expressing carcinogenicity. From the tens of descriptors calculated, principal component analysis enabled us to restrict the number of parameters to be used for the artificial neural network (ANN). We used 104 molecules for the study. An $R_{cv}^2 = 0.69$ was obtained. After removal of 12 outliers, a new ANN gave an $R_{cv}^2$ of 0.82.

## 1. INTRODUCTION

Man is exposed to many chemicals of natural and synthetic origin. An urgent question concerns their potential negative effects on human health. To identify chemicals inducing toxicity and to limit the incidence of human cancers and other diseases, rodent bioassays are the principal methods used today. However, this approach is not altogether problem-free, on several accounts: (1) the cost of the assay (>1 million U.S. dollars per chemical); (2) the time needed for the tests (3−5 years); (3) ethical considerations and public pressure to reduce or eliminate the use of animals in research and testing;[1] (4) difficulties in the extrapolation to man.

We were interested in the prediction of carcinogenicity, but cancer is not a single disease. Several mechanisms are involved in the various processes leading to the different tumors. This makes the task of assessing the computational prediction particularly challenging. Dedicated expert systems have been employed for computerized prediction of carcinogenicity.[2,3] However, these have limitations.[1−3] These expert systems work mainly on the assumption that toxicity is linked to the presence of toxic residues, either defined by human experts or found by the expert system. In some cases, the expert systems also use some simple physicochemical parameters. A very recent book describes the state-of-the-art of the research in the prediction of toxicity.[4]

Another widespread approach for predicting toxicity relies on molecular descriptors, which refer to global properties or characteristics of the molecule. In recent years a huge increase in the number of studies of theoretical molecular descriptors has appeared in the literature, including their use in toxicity prediction.[5] In the case of expert systems chemical data can be handled in several formats, but with artificial neural network (ANN) molecular descriptors are more suitable, and indeed they have been used in the prediction of carcinogenicity with contrasting results.[6−8] In this study we consider the use of molecular descriptors as input to ANN for the prediction of carcinogenicity of aromatic compounds with nitrogen-containing substituents.

## 2. METHODS

**2.1. Input and Output of the Model.** In many cases the carcinogenicity of a compound is classified by activity. A numerical, continuous approach was introduced by Gold and colleagues.[9] Gold's database contains standardized results for carcinogenicity for more than 1200 chemicals; for each substance it reports the carcinogenicity on rat and mouse, expressed using the parameter TD50, which is the chronic dose rate that would give half the animals tumors within some standard experimental time—the "standard lifespan" for the species. The huge amount of information in the database and the quantitative homogeneous evaluation are two important advantages. This database was therefore adopted as the basis for selecting the output parameter for the neural network. In the present study, for each chemical we chose the lowest (i.e. most potent) TD50. For the purpose of homogeneity all data refer to the mouse.

We limited the chemical compounds to be evaluated to those containing an aromatic ring and a nitrogen linked to the aromatic ring, because our previous experience with a commercial expert system showed that several of the compounds classified incorrectly belonged to this category.[10] The category includes several chemical classes, such as nitrosamines, amides, amines, and nitro derivatives, etc. The list of 104 selected compounds, with their toxic activity, is given in Table 1.

For the output we transformed the TD50 as follows: output = log(MW × 1000/TD50) (MW = molecular

10.1021/ci9903096 CCC: $18.00 
Published on Web 10/07/1999

**Table 1.** Chemical Names, CAS Number, and Experimental and Calculated Toxic Values of 104 Compounds

| name | CAS no. | expt | pred | name | CAS no. | expt | pred |
|---|---|---|---|---|---|---|---|
| (*N*-6)-(methylnitroso)adenine | | 0.6665 | 0.4923 | 5-nitroacenaphthene | 602-87-9 | 0.6194 | 0.6508 |
| (*N*-6)-methyladenine | 443-72-1 | 0.0000 | 0.4462 | acetaminophen | 103-90-2 | 0.0000 | 0.3215 |
| 1,5-naphthalenediamine | 2243-62-1 | 0.5838 | 0.5713 | AF-2 | 3688-53-7 | 0.5922 | 0.5664 |
| 1-(1-naphthyl)-2-thiourea | 86-88-4 | 0.8274 | 0.6979 | aniline·HCl | 142-04-1 | 0.2679 | 0.2523 |
| 1-amino-2-methylanthraquinone | 82-28-0 | 0.5516 | 0.6981 | anthranilic acid | 118-92-3 | 0.1737 | 0.1693 |
| 1-[(5-nitrofurfurylidene)amino]hydantoin | 67-20-9 | 0.4588 | 0.4831 | atrazine | 1912-24-9 | 0.6881 | 0.6902 |
| 2,2′,5,5′-tetrachlorobenzidine | 15721-02-5 | 0.5963 | 0.6738 | azobenzene | 103-33-3 | 0.7571 | 0.7360 |
| 2,2,2-trifluoro-*N*-[4-(5-nitro-2-furyl)-2-thiazolyl]acetamide | 42011-48-3 | 0.7321 | 0.6992 | benzidine·2HCl | 531-85-1 | 0.7086 | 0.6738 |
| | | | | c.i. disperse yellow 3 | 2832-40-8 | 0.4769 | 0.4644 |
| 2,4,5-trimethylaniline | 137-17-7 | 0.7129 | 0.6384 | chloramben | 133-90-4 | 0.3477 | 0.2602 |
| 2,4,6-trimethylaniline·HCl | 6334-11-8 | 0.6498 | 0.6310 | chlorambucil | 305-03-3 | 1.0000 | 0.9094 |
| 2,4-diaminoanisole sulfate | 39156-41-7 | 0.4965 | 0.4567 | cinnamyl anthranilate | 87-29-6 | 0.4017 | 0.4539 |
| 2,4-diaminotoluene·2HCl | 636-23-7 | 0.5643 | 0.5146 | d & c red no. 9 | 5160-02-1 | 0.4336 | 0.4384 |
| 2,4-dimethoxyaniline·HCl | 54150-69-5 | 0.4257 | 0.4197 | dacarbazine | 4342-03-4 | 0.8653 | 0.5674 |
| 2,4-dinitrophenol | 51-28-5 | 0.0000 | −0.0145 | dapsone | 80-08-0 | 0.0000 | 0.4293 |
| 2,4-dinitrotoluene | 121-14-2 | 0.0000 | 0.3873 | fd & c red no. 4 | 4548-53-2 | 0.2512 | 0.2209 |
| 2,4-xylidine·HCl | 21436-96-4 | 0.6608 | 0.5765 | fd & c yellow no. 6 | 2783-94-0 | 0.2717 | 0.2126 |
| 2,5-xylidine·HCl | 51786-53-9 | 0.4458 | 0.5227 | fluometuron | 2164-17-2 | 0.5344 | 0.4913 |
| 2,6-dichloro-*p*-phenylenediamine | 609-20-1 | 0.4405 | 0.4430 | formic acid 2-[4-(5-nitro-2-furyl)-2-thiazolyl]hydrazide | 3570-75-0 | 0.7277 | 0.6196 |
| 2-(acetylamino)fluorene | 53-96-3 | 0.7563 | 0.7638 | | | | |
| 2-amino-4-(5-nitro-2-furyl)thiazole | 38514-71-5 | 0.7243 | 0.6966 | furosemide | 54-31-9 | 0.4876 | 0.5560 |
| 2-amino-4-(*p*-nitrophenyl)thiazole | 2104-09-8 | 0.7133 | 0.6690 | hydrochlorothiazide | 58-93-5 | 0.4514 | 0.5654 |
| 2-amino-4-nitrophenol | 99-57-0 | 0.4384 | 0.4929 | *m*-cresidine | 102-50-1 | 0.5100 | 0.5057 |
| 2-amino-5-nitrophenol | 121-88-0 | 0.3238 | 0.3026 | *m*-phenylenediamine·2HCl | 541-69-5 | 0.4844 | 0.4144 |
| 2-amino-5-nitrothiazole | 121-66-4 | 0.0000 | −0.0862 | *m*-toluidine·HCl | 638-03-9 | 0.3831 | 0.3642 |
| 2-aminoanthraquinone | 117-79-3 | 0.4630 | 0.6501 | melamine | 108-78-1 | 0.3532 | 0.4286 |
| 2-aminodiphenylene oxide | 3693-22-9 | 0.7344 | 0.7324 | melphalan | 148-82-3 | 0.9803 | 1.0032 |
| 2-biphenylamine·HCl | 2185-92-4 | 0.4241 | 0.3075 | methotrexate | 59-05-2 | 0.6443 | 0.4927 |
| 2-chloro-*p*-phenylenediamine sulfate | 61702-44-1 | 0.4001 | 0.4022 | metronidazole | 443-48-1 | 0.4927 | 0.4924 |
| 2-hydrazino-4-(5-nitro-2-furyl)thiazole | 26049-68-3 | 0.6857 | 0.6391 | mexacarbate | 315-18-4 | 0.8264 | 0.8305 |
| 2-hydrazino-4-(*p*-aminophenyl)thiazole | 26049-71-8 | 0.7018 | 0.6003 | *N*-(l-naphthyl)ethylenediamine·2HCl | 1465-25-4 | 0.0000 | 0.2226 |
| 2-hydrazino-4-(*p*-nitrophenyl)thiazole | 26049-70-7 | 0.7134 | 0.6021 | *N*-nitrosodiphenylamine | 86-30-6 | 0.4952 | 0.4837 |
| 2-methyl-1-nitroanthraquinone | 129-15-7 | 0.8404 | 0.7969 | *N*-phenyl-*p*-phenylenediamine·HCl | 2198-59-6 | 0.0000 | 0.4836 |
| 2-naphthylamine | 91-59-8 | 0.6557 | 0.6456 | *N*-[4-(5-nitro-2-furyl)-2-thiazolyl]-formamide | 24554-26-5 | 0.7325 | 0.7051 |
| 2-nitro-*p*-phenylenediamine | 5307-14-2 | 0.4532 | 0.2208 | | | | |
| 2-*sec*-butyl-4,6-dinitrophenol | 88-85-7 | 0.8360 | 0.8256 | *N*-[5-(5-nitro-2-furyl)-1,3,4-thiadiazol-2-yl]acetamide | 2578-75-8 | 0.7440 | 0.5990 |
| 3,3′-dimethoxybenzidine-4,4′-diisocyanate | 91-93-0 | 0.2791 | 0.4109 | | | | |
| 3-(3,4-dichlorophenyl)-1,1-dimethylurea | 330-54-1 | 0.4788 | 0.6364 | nithiazide | 139-94-6 | 0.4609 | 0.4735 |
| 3-chloro-*p*-toluidine | 95-74-9 | 0.3807 | 0.3849 | nitrofen | 1836-75-5 | 0.6198 | 0.5780 |
| 3-nitro-*p*-acetophenetide | 1777-84-0 | 0.3995 | 0.4186 | *o*-aminoazotoluene | 97-56-3 | 0.5936 | 0.5913 |
| 4′-fluoro-4-aminodiphenyl | 324-93-6 | 0.8306 | 0.5675 | *o*-anisidine·HCl | 134-29-2 | 0.4162 | 0.4160 |
| 4,4′-methylenebis(2-chloroaniline)·2HCl | 64049-29-2 | 0.6141 | 0.6300 | *o*-phenylenediamine·2HCl | 615-28-1 | 0.4333 | 0.4379 |
| 4,4′-methylenebis(*N*,*N*-dimethyl)benzenamine | 101-61-1 | 0.5456 | 0.5662 | *o*-toluidine·HCl | 636-21-5 | 0.4296 | 0.4531 |
| 4,4′-methylenedianiline·2HCl | 13552-44-8 | 0.6760 | 0.6068 | *p*-anisidine·HCl | 20265-97-8 | 0.0000 | 0.3522 |
| 4,4′-oxydianiline | 101-80-4 | 0.6680 | 0.5299 | *p*-chloroaniline | 106-47-8 | 0.3917 | 0.3951 |
| 4-amino-2-nitrophenol | 119-34-6 | 0.0000 | 0.2578 | *p*-cresidine | 120-71-8 | 0.5986 | 0.5234 |
| 4-aminodiphenyl | 92-67-1 | 0.8312 | 0.6604 | *p*-isopropoxydiphenylamine | 101-73-5 | 0.4703 | 0.4558 |
| 4-chloro-*m*-phenylenediamine | 5131-60-2 | 0.4088 | 0.3924 | *p*-nitrosodiphenylamine | 156-10-5 | 0.5024 | 0.6000 |
| 4-chloro-*o*-phenylenediamine | 95-83-0 | 0.4233 | 0.4286 | *p*-phenylenediamine·2HCl | 624-18-0 | 0.3813 | 0.3249 |
| 4-chloro-*o*-toluidine·HCl | 3165-93-3 | 0.6942 | 0.6084 | pentachloronitrobenzene | 82-68-8 | 0.6161 | 0.6816 |
| 4-nitro-*o*-phenylenediamine | 99-56-9 | 0.0000 | 0.3094 | phenacetin | 62-44-2 | 0.2859 | 0.3255 |
| 4-nitroanthranilic acid | 619-17-0 | 0.2882 | 0.3812 | phenylhydrazine | 100-63-0 | 0.0000 | 0.0428 |
| 5-nitro-2-furaldehyde semicarbazone | 59-87-0 | 0.6600 | 0.4923 | proflavine·HCl hemihydrate | 952-23-8 | 0.6535 | 0.6667 |
| 5-nitro-*o*-anisidine | 99-59-2 | 0.4276 | 0.4530 | pyrimethamine | 58-14-0 | 0.5199 | 0.6243 |

weight), in order to have a more continuous output space and to refer to the moles of the chemical, not the weight.[8]

**2.2. Molecular Descriptors.** Chemical structures were drawn with Hyperchem (Hypercube, Inc.) and optimized using the PM3 Hamiltonian. We used the following programs to calculate descriptors: VAMP version 6.1 (Oxford Molecular Ltd.) for the quantum-mechanical and thermodynamic calculations, on a Silicon Graphics XS24 workstation; HAZARD EXPERT version 3.0 (CompuDrug Chemistry Ltd., Budapest, Hungary) for log *D* calculation; TSAR version 3.0 (Oxford Molecular) for the other descriptors, using a personal computer.

We calculated the 34 descriptors listed in Table 2. log *D* was calculated at pH 2, 7.4, and 10 as representative of the pH of the stomach, blood, and gut, where different processes may occur with the chemicals. The complete set of values is available from the authors on request.

**Table 2.** The 34 Used Descriptors

| | |
|---|---|
| molecular weight | three principal axes of inertia |
| log *D* at pH 2, 7.4, 10 | Balaban Index |
| HOMO | Wiener Index |
| LUMO | Randic Index |
| heat of formation | five Kier & Hall connectivity indices |
| dipole moment | |
| polarizability | six Kier shape indices |
| total energy | flexibility index |
| molecular volume | ellipsoidal volume |
| three principal moments of inertia | electrotopological sum |

**2.3. Reducing the Number of Descriptors by Principal Component Analysis.** We used principal component analysis

(PCA) to select a smaller set of descriptors so the network could converge faster.

The main change in the set of 104 molecules (accounting for 63% of the total variability) was explained by the descriptor total energy and by a pool of descriptors including topological, geometric, and electrostatic values inversely correlated with the first principal component (PC). The second PC, accounting for another 8% of the variability, was mainly related to the dipole moment, the topological index of Balaban, and the quantum-chemical HOMO and LUMO descriptors and to log *D* at pH 7.4 and pH 10. The log *D* at pH 2 correlated with the third PC, thus explaining another smaller but different source of variability.

Descriptors with the highest scores on the first four components of PCA (accounting for 85% of the total variability) were chosen and reduced, eliminating those most closely correlated. A final criterion was to keep a pool of descriptors representing the different aspects of the molecule considered (physicochemical, electronic, and topological, etc). From the 34 descriptors calculated, 13 were selected: molecular weight, HOMO, LUMO, dipole moment, polarizability, Balaban, ChiV3 and flexibility indices, log *D* at pH 2 and pH 10, third principal axis of inertia, ellipsoidal volume, and electrotopological sum.

**2.4. Artificial Neural Networks.** In all the simulations, performed with MBP v 1.1,[11] the working parameters were set as follows: the weight initialized with the SCAWI technique; net gain $\eta(0) = 0.75$; initial moment $\alpha(0) = 0.9$; acceleration factor YPROP, $K_a = 0.7$, $K_d = 0.07$. The algorithm stopped itself when it encountered one of the following conditions: gradient lower than $10^{-6}$; mean square error in validation (MSE) equal to 0; maximum calculated difference between calculated and desired output equal to 0; maximum number of iterations reached. Each network was trained starting from 100 random points in space, in order to minimize the probability of converging toward local minima. Input data were scaled between 0 and 1 in order to have a homogeneous range of variation of descriptors. The output was scaled accordingly.

For the validation step the leave-two-out approach was adopted, i.e. a cross-validation procedure using two examples in validation and the others for training. Five ANN models were generated, using data sets composed of 84 molecules randomly chosen in the training set and 20 in the test set.

The software is available on request, for noncommercial use.

## 3. RESULTS AND DISCUSSION

Most QSAR studies consider a limited number of parameters, taking account of previous knowledge in the field and using multivariate linear analysis. In our case there was no previous knowledge on the importance of specific molecular descriptors. We therefore considered a wide range of different classes, as detailed above, to extract information without a priori elimination of any possibilities.

We tried using regression analysis, but without success. ANN can be used to model complex phenomena where noise and nonlinear processes may be present, such as in our case. A disadvantage is the time needed, because many iterations are needed. This is a weakness of this neural network if we want to keep all the descriptors as inputs. Reducing the inputs shortens the training time. If this involves eliminating redundancy, the net has more chance of finding relevant parameters. For these reasons (reduction of computation time and elimination of redundancy) we chose PCA to select the ANN inputs, as it has been used for this purpose in other cases.[12,13] A risk related to the use of PCA is the possibility of eliminating inputs which behave nonlinearly. To verify that we had not eliminated any useful information, we built a new ANN using the first 12 PCs as inputs. These contain about 99% of the information of the original set of variables. The results with these PCs were comparable to those with the selected descriptors, shown below, indicating that we had not lost information through our selection.

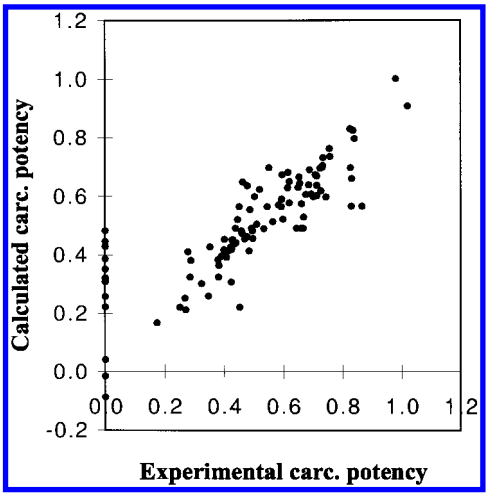


**Figure 1.** Predicted versus experimental carcinogenicity values with the BPNN four-neuron model.

**Table 3.** Results with BPNN of Increasing Numbers of Hidden Neurons (MSE = Mean Square Error)[a]

| neurons | MSE | $R_{cv}^2$ | neurons | MSE | $R_{cv}^2$ |
|---|---|---|---|---|---|
| 3 | 0.0157 | 0.675 | 6 | 0.0153 | 0.676 |
| 4 | *0.0146* | *0.691* | 7 | *0.0146* | *0.691* |
| 5 | 0.0154 | 0.676 | | | |

[a] Best results are in italics.

Table 1 gives results of the back-propagation neural network (BPNN). Figure 1 shows the predicted and experimental carcinogenicity with the BPNN four-neuron model.

The average $R^2$ cross-validated ($R_{cv}^2$) after 10 000 iterations and using different numbers of internal neurons is shown in Table 3.

To overcome possible representation bias in our data set, we built up five random data sets composed of 84 molecules in the training set and 20 in the test set. Then five independent ANN models were generated. This approach has been used recently.[14] $R_{cv}^2$ for these models was 0.70 using four or six neurons in the inner layer, in agreement with the leave-two-out method (see Table 3).

For the BPNN, the presence of outliers in the set was assumed and investigated in order to see whether the network's capacity for generalization improved after removing them and to assess the chemical nature of the activity of the compounds.

We adopted a conservative approach to remove outliers, taking only the molecules presenting an error in validation higher than 0.2 in the two best models (those with four and

**Table 4.** Results with BPNN of Increasing Number of Hidden Neurons, after Removing the Outliers (MSE = Mean Square Error)[a]

| neurons | MSE | $R_{cv}^2$ | neurons | MSE | $R_{cv}^2$ |
|---|---|---|---|---|---|
| 3 | 0.0062 | 0.793 | 6 | 0.0057 | 0.810 |
| 4 | *0.0053* | *0.824* | 7 | 0.0061 | 0.792 |
| 5 | 0.0053 | 0.824 | 8 | 0.0073 | 0.755 |

[a] Best results are in italics.

seven internal neurons). Twelve molecules were identified as outliers and removed. The results are presented in Table 4.

The results show that $R_{cv}^2$ has been clearly improved. Most of the outliers (9 out of 12) are molecules for which the experimental results for carcinogenicity were not statistically significant and an arbitrary value of $10^{31}$ was given in the Gold database (see Figure 1; they lie on the *y* axis, because of the transformation formula described in section 2.1 and scaling). The main experimental evidence for these molecules suggests noncarcinogenicity. Other considerations on the outliers regard their homogeneity from a chemical point of view. As we said, the compounds used for this ANN belong to several chemical classes and the outliers appear to be distributed over various chemical classes. Some are chemicals that have no structures in common with other members of the set, and this may explain their behavior. However, the ANN correctly predicted the toxicity of other chemicals which appear badly represented.

Special consideration must be given to two molecules, *o*- and *p*-anisidine. These isomers have identical or very similar chemical descriptors. However, their toxicity is very different, due to different metabolism in the animals. The ANN based on molecular descriptors was not able to distinguish them. This is a case of interesting behavior, shared with other compounds, which may undergo a metabolic process able to detoxify the chemical. In another study we solved the case of *o*- and *p*-anisidine by an expert system which distinguishes the toxic substructure.[15]

The present study illustrates the possibilities and limitations of the approach based on molecular descriptors. From the chemical point of view *o*- and *p*-anisidine may appear very similar, but for a living organism they are not. There are, however, chemicals which appear different within various chemical classes—as in the case of the compounds we have used—that the organism considers similar, because they are converted to aromatic amines. Knowledge of the body's bioprocesses is therefore an important source of information. Knowledge of the structural features of the molecule that characterize its specific mechanism of action cannot be ignored in some cases, in order to solve problems occurring in the prediction.

Another general point is the reliability of the database. We used an authoritative database, resulting from critical assessment of data from two sources: reports in the literature using different experimental protocols and results obtained according to a uniform protocol within the U.S. National Toxicology Program. Differences in the sources may affect the homogeneity of the data.[16] Furthermore, this database, like many others, changes constantly as new studies appear, adding knowledge.

A final comment on the database is that in most cases it still contains a limited number of compounds (despite the huge amount of work needed to build them up), so for some compounds we did not have enough examples to train the ANN properly.

## 4. CONCLUSIONS

Many models for toxicity prediction use linear relationships, which apply well within congeneric chemical classes. ANN has been used in limited cases. Villemin et al. used ANN to model polycyclic aromatic compounds in carcinogenic classes, obtaining good results.[6] Vracko obtained an *r* of 0.83, after removing the outliers, for a set of aromatic compounds belonging to different chemical classes.[8] Benigni and Richard, in a study using 280 compounds of various kinds, concluded that BPNN models fitted the training sets but had no general applicability.[7] The main feature of their study is the large differences between the structures of the molecules, much wider than in the other ANN used to predict carcinogenicity, including our present study.

The present study shows the feasibility of an ANN for predicting carcinogenicity of chemicals of various types. Several chemical classes are in fact present.

Our study attempts to illustrate how knowledge can be improved using ANN, probably because it is modeling nonlinearity. With chemical descriptors as input ANN is useful for cases where multilinear regression fails. We are aware of the limitations of this approach, which are common to other methods, as discussed. However, we believe that no single approach can cope with the vast problem of predictive toxicology, as already noted by other authors.[17] The next task is the extension to a wider set of chemicals. How to extract rules from the ANN is a major topic, and how to integrate ANN results with those from independent sources. We have already evaluated this last point in some cases, coupling expert systems and ANN within hybrid systems able to incorporate the best elements from each of the approaches.[15,18]

## ACKNOWLEDGMENT

We acknowledge the financial support of the European Commission (Grant ERB-CP94 1029 until 1998 and Grant ENV4-CT97-0508 since 1998) and NATO (Grant CRG 971505), from 1998. We thank Dr. Y.-t. Woo for useful discussion.

## REFERENCES AND NOTES

(1) Omenn, G. S. Assessing the Risk Assessment Paradigm. *Toxicology* **1995**, *102*, 23−28.
(2) Benfenati, E.; Gini, G. Computational Predictive Programs (Expert Systems) in Toxicology. *Toxicology* **1997**, *119*, 213−225.
(3) Dearden, J. C.; Barratt, M. D.; Benigni, R.; Bristol, D. W.; Combes, R. D.; Cronin, M. T. D.; Judson, P. N.; Payne, M. P.; Richard, A. M.; Tichy, M.; Worth, A. P.; Yourick, J. J. The Development and Validation of Expert Systems for Predicting Toxicity. *ATLA* **1997**, *25*, 223−252.
(4) Gini, G. C.; Katritzky, A. R. *Predictive Toxicology of Chemicals: Experiences and Impact of AI Tools*; AAAI Press: Menlo Park, CA, 1999; pp 1−155.
(5) Karelson, M.; Maran, U.; Wang, Y.; Katritzky, A. R. QSPR and QSAR Models Derived with CODESSA Multipurpose Statistical Analysis Software. In *Predictive Toxicology of Chemicals: Experiences and Impact of AI Tools*; *AAAI 1999 Spring Symposium Series*; Gini, G. C., Katritzky, A. R., Eds.; AAAI Press: Menlo Park, CA, 1999; pp 45−48.
(6) Villemin, D.; Cherqaoui, D.; Mesbash, A. Predicting Carcinogenicity of Polycyclic Aromatic Hydrocarbons from Back-Propagation Neural Network. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 1288−1293.

(7) Benigni, R.; Richard, A. M. QSARS of Mutagens and Carcinogens: Two Case Studies Illustrating Problems in the Construction of Models for Noncongeneric Chemicals. *Mutat. Res.* **1996**, *371*, 29−46.
(8) Vracko, M. A Study of Structure−Carcinogenic Potency Relationship with Artificial Neural Networks. The Using of Descriptors Related to Geometrical and Electronic Structures. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 1037−1043.
(9) Gold, L. S.; Slone, T. H.; Manley, N. B.; Backman Garfinkel, G.; Hudes, E. S.; Rohrbach, L.; Ames, B. N. The Carcinogenic Potency Database: Analyses of 4000 Chronic Animal Cancer Experiments Published in the General Literature and by the U.S. National Cancer Institute/National Toxicology Program. *Environ. Health Perspect.* **1991**, *96*, 11−15.
(10) Benfenati, E.; Tichy, M.; Malvè, L.; Grasso, P.; Gini, G. Expert Systems for Toxicity Prediction Based on Fragment Recognition: Evaluation of a Commercial System and Improved Approaches; American Chemical Society Meeting, Las Vegas, NV, Sep 8−12, 1997; Abstracts of paper, COMP 136.
(11) Anguita, D. *Matrix Back Propagation v 1.1: User's Manual*; 1993. Available through anonymous ftp at risc6000.dibe.unige.it.
(12) Miao, X.; Azimi-Sadjadi, M. R.; Tina, B.; Dubey, A. C.; Witherspoon, N. H. Detection of Mines and Minelike Targets Using Principal Component and Neural-Network Methods. *IEEE Trans. Neural Networks* **1998**, *9*, 454−463.
(13) Ventura, S.; Silva, M.; Pérez-Bendito, D.; Hervás, C. Computational Neural Networks in Conjunction with Principal Component Analysis for Resolving Highly Nonlinear Kinetics. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 287−291.
(14) Sussman, N. B.; Macina, O. T.; Claycamp, H. G.; Grant, S. G.; Rosenkranz, H. S. The Utility of Multiple Random Sampling in the Development of SAR Models. In *Predictive Toxicology of Chemicals: Experiences and Impact of AI Tools*; *AAAI 1999 Spring Symposium Series*; Gini, G. C., Katritzky, A. R., Eds.; AAAI Press: Menlo Park, CA, 1999; pp 45−48.
(15) Gini, G.; Lorenzini, M.; Vittore, A.; Benfenati, E.; Grasso, P. Some Results for the Prediction of Carcinogenicity Using Hybrid Systems. In *Predictive Toxicology of Chemicals: Experiences and Impact of AI Tools*; *AAAI 1999 Spring Symposium Series*; Gini, G. C., Katritzky, A. R., Eds.; AAAI Press: Menlo Park, CA, 1999; pp 139−143.
(16) Helma, C.; Gottmann, E.; Kramer, S.; Pfahringer, B. Data Quality Issues in Toxicological Knowledge Discovery. In *Predictive Toxicology of Chemicals: Experiences and Impact of AI Tools*; *AAAI 1999 Spring Symposium Series*; Gini, G. C., Katritzky, A. R., Eds.; AAAI Press: Menlo Park, CA, 1999; pp 8−11.
(17) Bahler, D.; Bristol, D. W. Prediction of Chemical Carcinogenicity in Rodents by Machine Learning of Decision Trees and Rule Sets. In *Predictive Toxicology of Chemicals: Experiences and Impact of AI Tools*; *AAAI 1999 Spring Symposium Series*; Gini, G. C., Katritzky, A. R., Eds.; AAAI Press: Menlo Park, CA, 1999; pp 74−77.
(18) Gini, G.; Testaguzza, V.; Benfenati, E.; Todeschini, R. HyTEx (Hybrid Toxicology Expert System): Architecture and Implementation of a Multi-domain Hybrid Expert System for Toxicology. *Chemom. Intell. Lab. Syst.* **1998**, *43*, 135−145.

CI9903096