Neural networks studies: Quantitative structure–activity relationships of antifungal 1-[2-(substituted phenyl)allyl]imidazoles and related compounds

ARTICLE *in* CHEMOSPHERE · APRIL 2001

Impact Factor: 3.34 · DOI: 10.1016/S0045-6535(00)00111-9

5 AUTHORS, INCLUDING:

Souad Mghazli
Cadi Ayyad University
3 PUBLICATIONS **17** CITATIONS
SEE PROFILE

Didier Villemin
Université de Caen Normandie
293 PUBLICATIONS **2,868** CITATIONS
SEE PROFILE

Driss Cherqaoui
Cadi Ayyad University
38 PUBLICATIONS **447** CITATIONS
SEE PROFILE

PERGAMON

Chemosphere 43 (2001) 385–390

CHEMOSPHERE

Neural networks studies: quantitative structure–activity relationships of antifungal 1-[2-(substituted phenyl)allyl]imidazoles and related compounds

Souâd Mghazli [a], Abderrahim Jaouad [a], Mohamed Mansour [b,✠], Didier Villemin [c], Driss Cherqaoui [a,*]

[a] *Département de Chimie, Faculté des Sciences Semlalia, Université Cadi Ayyad, BP 2390 Marrakech, Morocco*
[b] *GSF-National Research Center for Environment and Health, Institue of Ecological Chemistry Ingolstadter Langstrasse 1, D-85764 Neuherberg, Germany*
[c] *Ecole Nationale Supérieure d'Ingénieurs (ENSI) ISMRA, LCMT, UMR CNRS no. 6507, 6 boulevard Maréchal Juin, 14050 Caen Cedex, France*

Received 30 June 1999; accepted 1 March 2000

Dr. Mohamed Mansour died last year in an automobile accident. This publication is dedicated to his memory by his co-authors

Abstract

Models of relationships between structure and antifungal activity of 1-[2-(substituted phenyl)allyl]imidazoles and related compounds were constructed by means of a multilayer neural network using the back-propagation (BP) algorithm. Each molecule was described by three structural and one physicochemical parameters. The leave-one-out procedure was used to assess the predictive ability of a neural network model. The results obtained were compared to those given in the literature by the multiple linear regression (MLR), and were found to be better. The contribution of each descriptor to the structure–activity relationships was evaluated. Hydrophobicity of the molecule was confirmed to take the most relevant part in the molecular description. © 2001 Elsevier Science Ltd. All rights reserved.

Keywords: Antifungal; Quantitative structure–activity relationships; Neural network; Back-propagation

1. Introduction

The study of relationships between molecular structures and antifungal activity has hastened the development of new active compounds. A certain number of computational techniques have been found useful for the establishment of these relationships (Hansch and Leo, 1995). A relatively recent technique and one that shows considerable promise is that of Neural Networks (NNs) (McClelland et al., 1988). NNs are artificial systems simulating the function of the human brain, where very high numbers of information-processing neurons are interconnected. They can handle problems involving imprecise or "noisy" data as well as problems that are highly non-linear and complex. NNs can identify and learn correlative patterns between sets of input data and corresponding target values. An NN must be trained by being repeatedly fed input data together with their corresponding target outputs. After a sufficient number of training iterations, the NN learns to recognize the relationship between input and output data and creates an internal model of the process governing the data. The NN can then use this internal model to make predictions for new inputs.

The application of NNs appeared in several areas of chemistry (Burns and Whitesides, 1993). NNs have been

* Corresponding author. Fax: +212-4-43-7408.
E-mail address: cherqaoui@ucam.ac.ma (D. Cherqaoui).

PII: S0045-6535(00)00111-9

applied to the identification of proton-NMR spectra (Thomsen and Meyer, 1989), the interpretation of IR spectra (Robb and Munk, 1990; Munk et al., 1991), the prediction of ^{13}C chemical shifts (Kvasnicka, 1991), the classification of mass spectra (Curry and Rumelhart, 1990), the estimation of aqueous solubilities (Bodor et al., 1991), the determination of protein structure (Qian and Sejnowski, 1988; Holley and Karplus, 1989), the investigation of quantitative structure–activity relationships (QSAR) (Adler et al., 1992; Domine et al., 1993; Villemin et al., 1993; Cherqaoui et al., 1998) and the prediction of chemical reactivity (Simon et al., 1993).

The goal of the current work is:

(a) to provide an application of NNs to the structure–antifungal activity relationships of 1-[2-(substituted phenyl)allyl]imidazoles and related compounds,

(b) to compare the results obtained by the NN to those given in the literature by multiple linear regression (MLR),

(c) to measure the contribution of each descriptor to structure–antifungal activity relationships.

2. Compounds studied and descriptors used

The set of 45 antifungal 1-[2-(substituted phenyl)allyl]imidazoles and related compounds used in the present paper had the same parent skeleton (Fig. 1). Those compounds and their antifungal activities have been studied by Kataoka et al. (1988) and are listed in Table 1. Each molecule was described by four molecular descriptors:

1. $\log P$ – molecular hydrophobicity,
2. V_w – van der Waals volume of substituent Y,
3. I_H takes a value of one for unsubstituted derivatives ($Y = H$) at the Y position,
4. I_{sat} takes one for the derivatives having no double bond at the β-carbon of the substituent on the imidazole ring.

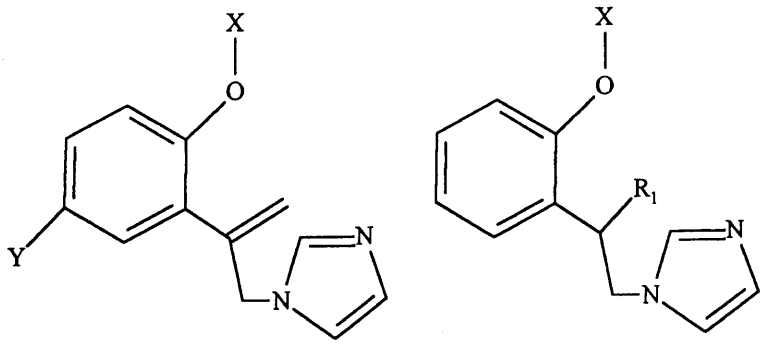
$\log P$, V_w, I_H and I_{sat} of the compounds studied are listed in Table 2.

3. Neural network

All the feed-forward NNs used in this paper are three-layer networks with four units ($\log P$, V_w, I_H and I_{sat}) in the input layer, a variable number of hidden neurons, and one unit (antifungal activity) in the output layer. A bias term was added to the input and hidden layers. Fig. 2 shows an example of the architecture of such an NN. Each neuron of the input layer is fully interconnected with each neuron of the hidden layer which in turn is fully interconnected with the output neuron. There are no connections between the neurons within a layer nor any direct connection between those of the input and output layers. Input and output data are normalized between 0.1 and 0.9. The sigmoidal transfer function used in the NN is given by Eq. (1). In this equation O_i and O_j are the outputs of neuron i and j, respectively, and W_{ij} is the weight connecting neuron i to neuron j.

$$O_i = \left(1 + \exp\left(-\Sigma W_{ij} O_j\right)\right)^{-1} \quad (1)$$

The weights of the connections between the neurons were initially assigned with random values uniformly distributed between −0.5 and +0.5 and no momentum was added. The back-propagation (BP) algorithm was used to adjust those weights. This algorithm has been described previously (Cherqaoui and Villemin, 1994) with a simple example of application and further details of this algorithm are given elsewhere (Freeman and



A B

Fig. 1. Common skeleton (A or B) of the compounds studied (X = various substituents; Y = H, Me, F, Cl; $R_1 = CH_3$, CH_3CH_2).

Table 1
Compounds studied (Fig. 1) and their antifungal activities[a]

Compound (A)	*X*	*Y*	Antifungal activity
Mol. 1	CH_2=$CHCH_2$	F	4.89
Mol. 2	CH_2=$CHCH_2$	Cl	5.16
Mol. 3	CH≡CCH_2	F	4.63
Mol. 4	CH≡CCH_2	Cl	4.67
Mol. 5	$CH_3(CH_2)_7$	CH_3	6.67
Mol. 6	$CH_3(CH_2)_7$	Cl	6.86
Mol. 7	Bz	CH_3	5.66
Mol. 8	Bz	F	5.81
Mol. 9	Bz	Cl	6.13
Mol. 10	3-Cl–Bz	CH_3	6.49
Mol. 11	3-Cl–Bz	F	6.56
Mol. 12	3-Cl–Bz	Cl	6.70
Mol. 13	2,4-Cl_2–Bz	CH_3	6.86
Mol. 14	2,4-Cl_2–Bz	F	7.19
Mol. 15	2,4-Cl_2–Bz	Cl	7.14
Mol. 16	2,6-Cl_2–Bz	F	6.72
Mol. 17	3,4-Cl_2–Bz	F	6.82
Mol. 18	3,4-Cl_2–Bz	Cl	6.99
Mol. 19	n-C_3H_7	H	4.65
Mol. 20	i-C_3H_7	H	5.09
Mol. 21	CH_2=$CHCH_2$	H	4.23
Mol. 22	CH=CCH_2	H	3.98
Mol. 23	$CH_3(CH_2)_3$	H	5.26
Mol. 24	$CH_3(CH_2)_7$	H	6.60
Mol. 25	$CH_3(CH_2)_9$	H	6.85
Mol. 26	Bz	H	5.00
Mol. 27	3-Cl–Bz	H	6.20
Mol. 28	4-Cl–Bz	H	5.95
Mol. 29	4-C_6H_5–Bz	H	7.26
Mol. 30	2,4-Cl_2–Bz	H	7.14
Mol. 31	3,4-Cl_2–Bz	H	6.81
Mol. 32	$C_6H_5CH(CH_3)$	H	5.39
Mol. 33	$C_6H_5O(CH_2)_2$	H	5.54
Mol. 34	$C_6H_5(CH_2)_2$	H	5.63
Mol. 35	$C_6H_5(CH_2)_4$	H	6.30
Mol. 36	$C_6H_5(CH_2)_6$	H	6.61

Compound (B)	*X*	R_1	Antifungal activity
Mol. 37	n-C_3H_7	CH_3	4.39
Mol. 38	CH_2=$CHCH_2$	CH_3	3.77
Mol. 39	Bz	CH_3	4.92
Mol. 40	2,4-Cl_2–Bz	CH_3	6.44
Mol. 41	n-C_3H_7	CH_3CH_2	4.30
Mol. 42	CH_2=$CHCH_2$	CH_3CH_2	3.84
Mol. 43	Bz	CH_3CH_2	4.79
Mol. 44	4-Cl–Bz	CH_3CH_2	5.75
Mol. 45	2,4-Cl_2–Bz	CH_3CH_2	6.40

[a] Symbol Bz indicates the benzyl group.

Skapura, 1991). The learning rate was initially set to one and was gradually decreased until the error function could no longer be minimized.

All calculations of NNs were done on 333 MHz PENTIUM computer using our program written in C language.

4. Results and discussion

Three different sessions has been achieved: computation, prediction and the descriptor's contribution. The first was aimed at checking the NN learning performance as well as the molecular descriptors adequacy.

Table 2
$\log P$, V_w, I_H and I_{sat} of the compounds studied

Compound	$\log P$	V_w	I_H	I_{sat}
Mol. 1	3.05	0.580	0	0
Mol. 2	3.62	1.165	0	0
Mol. 3	2.18	0.580	0	0
Mol. 4	2.75	1.165	0	0
Mol. 5	6.60	1.367	0	0
Mol. 6	6.80	1.165	0	0
Mol. 7	4.67	1.367	0	0
Mol. 8	4.30	0.580	0	0
Mol. 9	4.87	1.165	0	0
Mol. 10	5.38	1.367	0	0
Mol. 11	5.01	0.580	0	0
Mol. 12	5.58	1.165	0	0
Mol. 13	6.09	1.367	0	0
Mol. 14	5.73	0.580	0	0
Mol. 15	6.30	1.165	0	0
Mol. 16	5.73	0.580	0	0
Mol. 17	5.73	0.580	0	0
Mol. 18	6.30	1.165	0	0
Mol. 19	3.31	0.340	1	0
Mol. 20	3.31	0.340	1	0
Mol. 21	2.76	0.340	1	0
Mol. 22	1.90	0.340	1	0
Mol. 23	3.84	0.340	1	0
Mol. 24	5.95	0.340	1	0
Mol. 25	7.01	0.340	1	0
Mol. 26	4.02	0.340	1	0
Mol. 27	4.73	0.340	1	0
Mol. 28	4.73	0.340	1	0
Mol. 29	5.90	0.340	1	0
Mol. 30	5.44	0.340	1	0
Mol. 31	5.44	0.340	1	0
Mol. 32	4.33	0.340	1	0
Mol. 33	3.93	0.340	1	0
Mol. 34	4.35	0.340	1	0
Mol. 35	5.40	0.340	1	0
Mol. 36	6.46	0.340	1	0
Mol. 37	3.76	0.340	1	1
Mol. 38	3.22	0.340	1	1
Mol. 39	4.47	0.340	1	1
Mol. 40	5.89	0.340	1	1
Mol. 41	4.29	0.340	1	1
Mol. 42	3.74	0.340	1	1
Mol. 43	5.00	0.340	1	1
Mol. 44	5.71	0.340	1	1
Mol. 45	6.43	0.340	1	1

The second was aimed at determining the predictive ability of a trained NN. In the third session, we attempted an evaluation of the importance of the descriptors used.

4.1. Computation

In a BP NN the input and output neurons are known since they represent, respectively – in this study – the descriptors used and the antifungal activity. 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 (Zupan and Gasteiger, 1993). For the determination of the number of hidden neurons, we have recently (Cherqaoui and Villemin, 1994) discussed the usefulness of ρ parameter defined as

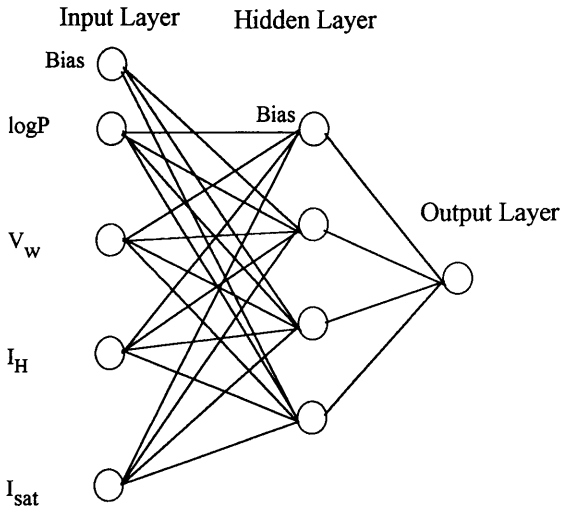


Fig. 2. Schematic representation of a three-layer neural network. The configuration shown is 4-3-1 (the bias unit is not included in the unit count).

$$\rho = \frac{\text{Number of data points in the training set}}{\text{Sum of the number of connections in the NN}}.$$

The range of $1 < \rho < 2.2$ has been suggested (Andrea and Kalayeh, 1991; Zupan and Gasteiger, 1993) as an empirical guideline of acceptable ρ values. It has been claimed that for $\rho \ll 1.0$ the NN simply memorizes the data. While for $\rho \gg 3.0$, the NN is not able to generalize. So we used a number of neurons in the hidden layer allowing to maintain ρ in the $1 < \rho < 3$ range, to avoid these two problems. Five different architectures have been applied (4-x-1; $x = 3,4,5,6,7$). Each architecture was trained with 10 different initial random sets of weight and with the number of cycles limited to 1000. In all cases 100 cycles were enough to obtain stable results. The criteria used for the comparison of the five architectures are the correlation coefficient (R) and the standard deviation of error (S) (Cherqaoui and Villemin, 1994).

All the results given by NNs and MLR are shown in Table 3. We see that in all the cases the NN approach gives better results than MLR. This preliminary study allows to conclude that all the NN architectures were

Table 3
Comparison of standard error of computation (S) and correlation coefficient (R) given by NN and MLR

NN's configuration	S	R^2
4-3-1	0.229	0.950
4-4-1	0.229	0.950
4-5-1	0.227	0.951
4-6-1	0.220	0.954
4-7-1	0.220	0.954
MLR	0.319	0.914

able to establish a satisfactory relationship between the molecular descriptors and the antifungal activity.

4.2. Prediction

After determining the range of hidden neurons giving a good computation, the most important predictive aspect of NNs was studied: the prediction of the antifungal activity of new molecules. To determine that predictive aspect, leave-one-out procedure has been used. In this procedure one compound is removed from the data set, the network is trained with the remaining compounds and used to predict the discarded compound. The process is repeated in turn for each compound in the data set. In the next step, the predictive ability of different networks was assessed by the standard error of prediction (SEP) and the leave-one-out R^2.

The computation time varied depending on the architecture of the NN and also on the size of the data set. In this study the average training time for each run (100 cycles) was about 30 s. Leave-one-out procedure was also used to assess the predictive power of the MLR. MLR was applied to the same data set and to the same four molecular descriptors.

The results obtained are presented in Table 4. They are satisfactory and confirmed that the four molecular descriptors are appropriate for the compounds studied.

The results given by all the architectures tried was more satisfactory than those of the MLR method. In MLR the relationship between antifungal activity and the four descriptors is expressed by a linear combination. On the contrast, one crucial aspect of the predictive performance of a NN used to solve computation problems is its non-linear power. Note that in the present work, MLR gave a SEP of 0.337, indicating that the function mapped by the NN is not so far from linear.

Table 4
Comparison of predictive ability for NNs and MLR using leave-one-out procedure

NN's configuration	SEP	R^2
4-3-1	0.270	0.930
4-4-1	0.280	0.926
4-5-1	0.273	0.929
4-6-1	0.277	0.927
4-7-1	0.290	0.922
MLR	0.337	0.892

Table 5
Contributions (C_i) of descriptors to the QSAR

Parameter	C_i (4-3-1)	C_i (4-4-1)	C_i (4-5-1)
$\log P$	65	64	64
I_H	13	13	13
I_{sat}	13	13	13
V_w	9	10	10

4.3. Contribution of descriptors to QSAR

The evaluation of the relevance of descriptors proved quite interesting and useful. That is why we chose to estimate their relative contribution. The contribution of descriptor i ($i=1$–4) was estimated from the trained 4-x-1 configuration network ($x=3,4,5$). The descriptor under study was removed from the 4-x-1 NN together with its corresponding weights. Then the network (3-x-1) calculated the output of each molecule as usual. The mean of the deviations' absolute values Δm_i, between the observed antifungal activities and the estimated ones for all compounds, was calculated. This process was reiterated for each descriptor. Finally, the contribution (C_i) of descriptor i is given by

$$C_i = 100 * \Delta m_i \Big/ \sum_{i=1}^{4} \Delta m_i. \quad (2)$$

Table 5 shows that the three architectures of NN had identical results. They indicate that the relative importance of the descriptors varied in the following order: $\log P > I_H = I_{sat} > V_w$. The descriptor related to hydrophobic properties seems to be very important in the establishment of the quantitative structure–antifungal activity relationships. This confirms the finding of the previous study according to which antifungal activity is related to hydrophobic effects (Kataoka et al., 1988).

In order to assess their importance, the first three descriptors ($\log P$, I_H and I_{sat}) was used in a 3-6-1 NN. Six hidden neurons were chosen so as to maintain ρ between 1 and 3. The leave-one-out procedure results obtained ($s=0.302$ and $r=0.913$ with 100 epochs) show the relevance of these three descriptors.

5. Conclusion

This paper has discussed the use of BP NN to predict the antifungal activity of 1-[2-(substituted phenyl)al-