



Exploring QSAR of Non-Nucleoside Reverse Transcriptase Inhibitors by Neural Networks: TIBO Derivatives

Latifa Douali,^{1*} Didier Villemin² and Driss Cherqaoui¹

¹Département de Chimie, Faculté des Sciences Semlalia BP 2390 Université Cadi Ayyad, Marrakech, Morocco. Tel: (+212) 44 43 46 49 ; Fax : (+212) 44 43 74 08; E-mail: l.douali@ucam.ac.ma; d.cherqaoui@menara.ma

²Ecole Nationale Supérieure d'Ingénieurs (E.N.S.I.) I. S. M. R. A., LCMT, UMR CNRS n° 6507, 6 boulevard Maréchal Juin, 14050 Caen Cedex, France.

*Author to whom correspondence should be addressed.

Received: 27 April 2003 / Accepted: 18 September 2003 / Published: 30 January 2004

Abstract: Human Immunodeficiency Virus type 1 (HIV-1) reverse transcriptase is an important target for chemotherapeutic agents against the AIDS disease. 4,5,6,7-Tetrahydro-5-methylimidazo[4,5,1-jk][1,4]benzodiazepin-2(1H)-ones (TIBO) derivatives are potent non-nucleoside reverse transcriptase inhibitors (NNRTIs). In the present work, quantitative structure-activity relationship (QSAR) analysis for a set of 82 TIBO derivatives has been investigated by means of a three-layered neural network (NN). It has been shown that NN can be a potential tool in the investigation of QSAR analysis compared with the models given in the literature. NN gave good statistical results both in fitting and prediction processes ($0.861 \leq r^2 \leq 0.928$, $0.839 \leq q^2 \leq 0.845$). The relevant factors controlling the anti-HIV-1 activity of TIBO derivatives have been identified. The results are along the same lines as those of our previous studies on HEPT derivatives and indicate the importance of the hydrophobic parameter in modeling the QSAR for TIBO derivatives.

Keywords: HIV-1, TIBO, QSAR, neural network.

Introduction

The treatment of the acquired immunodeficiency syndrome (AIDS) is the most challenging worldwide medical problem. Most of the current strategies for treating AIDS depend on inhibiting HIV-1 reverse transcriptase enzyme. In this context, the non-nucleoside reverse transcriptase inhibitors

(NNRTIs) gained the greatest importance because of their specificity and their low cytotoxicity [1]. NNRTIs now comprise a very large number of chemically diverse compounds including the 4,5,6,7-Tetrahydro-5-methylimidazo[4,5,1-jk][1,4]benzodiazepin-2(1H)-ones (TIBO) derivatives [2]. The general structure of TIBO derivatives is represented in Figure 1. The TIBO derivatives have been the aim of numerous quantitative structure-activity relationships (QSAR) studies [3-5].

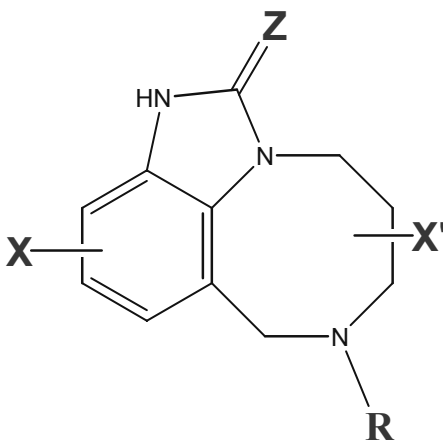


Figure 1. General structure of TIBO derivatives (X, X', Z and R: substituents).

In fact QSAR, as a technique attempting to summarize chemical and biological information in order to generate relationships between structure and biological activity, hastens the drug design and aims to develop these compounds.

A certain number of computational techniques have been found useful for the establishment of these QSAR such as MLR [4] and CoMFA [6]. A relatively recent technique and one that shows considerable promise is that of Neural Networks (NN) [7]. As far as we are concerned there are no reports on the use NN in the QSAR studies for the TIBO derivatives.

NN are artificial systems simulating the function of the human brain where very high numbers of information-processing neurons are interconnected. NN models are designed to recognize patterns between molecular descriptors (input data) and biological activity (target data) to produce forecasts (output data). The NN offer a number of advantages over the traditional statistical methods, due to their generalization, massive parallelism and ability to offer real time solutions. In addition, NN give good results when statistical techniques reach their limitations [8], especially in handling non-linearity in data sets [9].

The application of NN appeared in several areas of chemistry and biology [7]. NN have been applied also to the investigation of several QSAR [11-14].

The purpose of the current work is to provide an application of NN to the structure-anti-HIV-1 activity relationship of TIBO compounds. The results obtained by the NN will be compared to those given, in the literature, by multiple linear regression (MLR). Thereafter, we sought to measure the contribution of each descriptor to the structure-anti-HIV-1 activity relationship.

Materials and Methods

Compounds Studied

A series of 82 TIBO compounds [4] were taken under consideration in this study. All the molecules studied had the same parent skeleton (Figure 1). The structures and anti-HIV-1 activities of these compounds were described previously [4]. The anti-HIV activity of the compounds has been expressed by the compound's ability to protect MT-4 cells against the cytopathic effect of the virus. The concentration of the compound leading to 50% effect has been measured and expressed as IC_{50} . The logarithm of the inverse of this parameter has been used as biological end points ($\log 1/C$) in the QSAR studies.

Molecular Descriptors Used

In back-propagation NN, the input layer contains information concerning the data samples under study. In chemistry and biology, this information is represented by molecular descriptors. In our study, each molecule was described by 4 descriptors, which are given by Garg et al. [4]. These descriptors characterize the hydrophobic, the steric and the electronic aspects, respectively:

ClogP (or logP) : the calculated octanol/water partition coefficient of the molecule

$B_{1(8-x)}$: Verloop's sterimol parameter (width parameter of the X substituent at the position 8)

$I_R = 1$ if R = 3,3-dimethylallyl and $I_R = 0$ for others (see Fig. 1)

$I_Z = 1$ if Z = Oxygen and $I_Z = 0$ if Z = Sulphur (see Fig. 1)

Neural Network

All the feed-forward NN used in this paper are three-layer networks with four units (ClogP, $B_{1(8-x)}$, I_Z , I_R) in the input layer, a variable number of hidden neurons, and one unit ($\log 1/C$) in the output layer. A bias term was added to the input and hidden layers. Figure 2 shows an example of the architecture of such NN. Each neuron in any layer is fully connected with the neurons of a succeeding layer. There are neither connections between the neurons within a layer nor any direct connection between those of the input and the output layers. Input and output data are normalized between 0.1 and 0.9. The sigmoid function was used as the transformation function [7]. 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 algorithm was used to adjust those weights. This algorithm has been described previously [15] with a simple example of application and a detail of this algorithm is given elsewhere [7]. The learning rate was initially set to 1 and was gradually decreased until the error function could no longer be minimized.

All calculations of NN were done on 1.7 MHz Pentium 4 computer using our program written in C language.

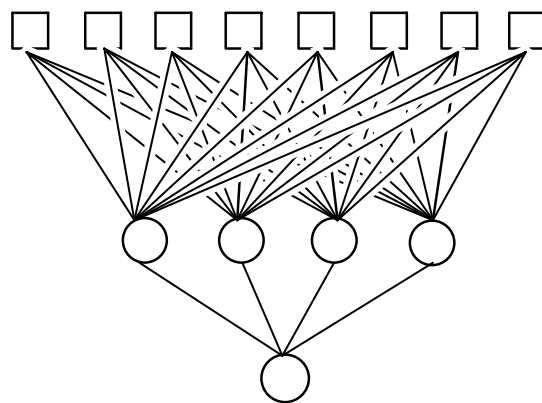


Figure 2. 8-4-1 architecture of a three-layer neural network.

Results and Discussion

In this work, computation, prediction and the descriptor's contribution are considered. The first was aimed at checking the NN learning performance as well as the molecular descriptors adequacy. The second was aimed at determining the predictive ability of a trained NN. In the third session, we attempt to evaluate the importance of the descriptors used.

Computation

In a back-propagation NN the input and output neurons are known since they represent respectively, in this study, the descriptors used and the anti-HIV-1 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 NN to chemistry, one hidden layer seems to be sufficient [10]. For the determination of the number of hidden neurons, some authors [16,17] have proposed a parameter ρ , which plays a major role in determining the best NN architecture. ρ is defined as follows:

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

In order to avoid overfitting it is recommended that $1 < \rho < 2.2$ [17]. 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. Zupan and Gasteiger [10] suggested that the number of connections in the network should be less than the number of points data.

We have then varied the number of neurons in the hidden layer to maintain ρ in the $1 < \rho < 3$ range. The learning performance of NN depends on the number of iterations but sufficient convergence was usually obtained after 10000 iterations. The results are reported in Table 1. In that same table the results achieved by Garg et al. [4] using the MLR method are given.

As it can be seen from Table 1, high correlation coefficients have been obtained by means of NN. It is noteworthy that all architectures tried give approximately the same value. In addition, we noticed

Table 1. Standard error of computation (SEC) and correlation coefficient (R^2) obtained by NN trained with 82 data points.

Architecture	R^2	SEC
4-5-1	0.910	0.432
4-6-1	0.924	0.395
4-7-1	0.925	0.394
4-8-1	0.924	0.395
4-9-1	0.923	0.399
4-10-1	0.922	0.401
4-11-1	0.927	0.388
4-12-1	0.923	0.399
4-13-1	0.928	0.387
MLR [4]	0.861	0.550

that in all cases the NN approach gives better results than MLR. The standard errors of calculation are lower and the correlation coefficients are higher with NN than with regression analysis. This preliminary study enables us to conclude that all the NN architectures were able to establish a satisfactory relationship between the molecular descriptors and the anti-HIV-1 activity.

Prediction

The predictive ability of an NN is its ability to give a satisfying output to a molecule not included in the examples the NN learned. 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. After leave-one-out procedure, the predictive ability of different networks was assessed by the standard error of prediction (SEP) and the leave-one-out R^2 [9]. The results of this analysis are shown in Table 2. These results are satisfying and show that the NN give correct predictions. All the NN architectures and MLR method give practically the same results ($R^2=0.840$ and $s = 0.575$). These results indicate that the function mapped by the NN is not so far from linear. The NN were able to extract information from samples to develop an internal representation of the anti-HIV-1 activity of TIBO without explicitly incorporating rules into the network.

Descriptor's Contribution

One of the purposes of QSAR analyses is to understand the forces governing the activity of a particular class of compounds and to assist drug design. Therefore, the evaluation of the descriptors relevance proved quite interesting and useful to shed more light on the structure-anti-HIV-1 activity. That is why we choose to estimate their relative contribution. The contribution of each descriptor was estimated from the trained 4-x-1-configuration network (x=5-13) using a technique proposed by Cherqaoui et al. [15]. All architectures of NN had identical results (Figure 3).

Table 2. Leave-one-out R^2 and standard error of prediction (SEP) obtained by NN trained with 82 data points.

Architecture	R^2	SEP
4-5-1	0.834	0.586
4-6-1	0.844	0.567
4-7-1	0.845	0.565
4-8-1	0.842	0.571
4-9-1	0.839	0.576
4-10-1	0.841	0.573
4-11-1	0.840	0.575
4-12-1	0.843	0.570
4-13-1	0.841	0.572

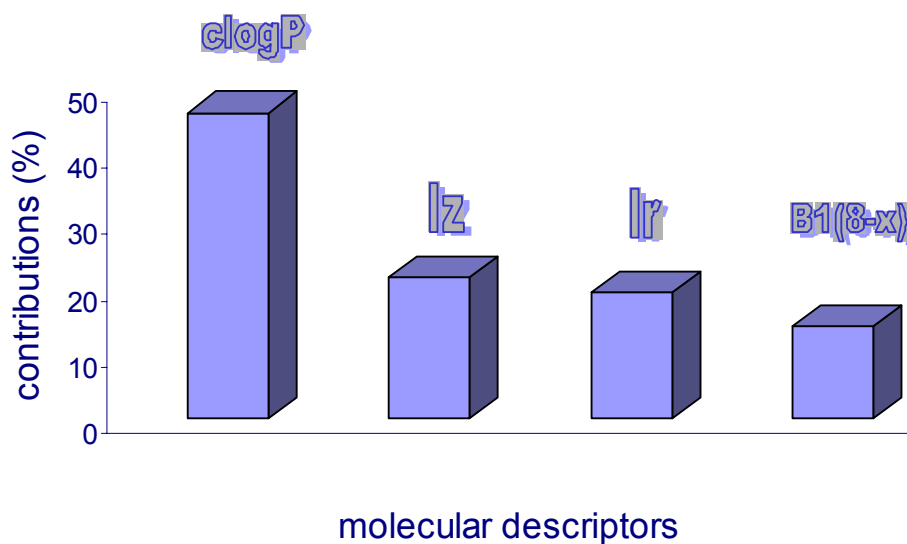
**Figure 3.** Contributions of descriptors to the QSAR.

Figure 3 indicates that the relative importance of the descriptors varied in the following order: $ClogP > I_Z > I_R > B1_{(8-x)}$. We can notice that the descriptor related to the hydrophobic property is the most important in the establishment of the QSAR of TIBO derivatives. This confirms the findings of the previous study according to which anti-HIV-1 activity is related to hydrophobic effect [4] and is along the same lines as those of our previous studies on HEPT derivatives [8]. Descriptors I_Z , I_R and $B1_{8-x}$ seem to be important in the establishment of the structure-ant-HIV-1 activity relationships. So, the inhibitory activity of TIBO is also governed by electronic (I_Z) and steric effects (I_R and $B1_{(8-x)}$).

Conclusion

A back-propagation neural network was applied to analyze the QSAR of TIBO compounds. The results obtained show that the NN was able to establish a satisfactory relationship between the molecular descriptors and the anti-HIV activity. The NN approach would seem to have great potential for determining quantitative structure-anti-HIV-1 activity relationships and as such be a valuable tool for the chemist. The main factor controlling the anti-HIV activity of TIBO derivatives have been determined by NN. Hydrophobicity of the compounds was thus found to take the most relevant part in the molecular description.

References

1. De Clercq, E. Perspective of Non-Nucleoside Reverse Transcriptase Inhibitors (NNRTIs) in the Therapy of HIV-1 Infection. *Il Farmaco* **1999**, *54*, 26-45.
2. Pauwels, R.; Andries, K.; Desmyter, J.; Schols, D.; Kukla, M. J.; Breslin, H. J.; Raeymaechers, A.; Gelder, J. V.; Woestenborgs, R.; Heykants, J.; Schellekens, K.; Janssen, M. A. C.; De Clercq, E.; Janssen, P. A. J. Potent and Selective Inhibition of HIV-1 Replication in vitro by a Novel Series of TIBO Derivatives. *Nature* **1990**, *343*, 470-474.
3. Hannongbua, S.; Pungpo, P.; Limtrakul, J.; Wolschann, P. Quantitative Structure-Activity Relationships and Comparative Molecular Field Analysis of TIBO Derivatives HIV-1 Reverse Transcriptase Inhibitors. *J. Comput.-Aided. Mol. Des.* **1999**, *7*, 2283-2292.
4. Garg, R.; Gupta, S. P.; Gao, H.; Babu, M. S.; Debnath, A. K. Comparative Quantitative Structure-Activity Relationship Studies on Anti-HIV Drugs. *Chem. Rev.* **1999**, *99*, 3525-3601.
5. Huuskonen, J. QSAR Modeling with the Electrotopological State: TIBO Derivatives. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 425-429.
6. Barreca, M. L.; carotti, A.; Carrieri, A.; Chimirri, A.; Monforte, A. M.; pellegrini Calace, M.; Rao, A. Comparative Molecular Field Analysis (CoMFA) and Doking Studies of Non-Nucleoside HIV-1 RT Inhibitors (NNIs). *Bioorg. Med. Chem.* **1999**, *7*, 2283-2292.
7. Freeman, J. A.; Skapura, D. M. *Neural Networks Algorithms, Applications, and programming Techniques*; Addition Wesley Publishing Company: Reading, 1991; pp 89-125.
8. Douali, L.; Villemin, D.; Cherqaoui, D. Comparative QSAR based on Neural Networks for the Anti-HIV Activity of HEPT Derivatives. *Curr. Pharm. Des.* **2003**, *9*, 1817-1826.
9. Douali, L.; Villemin, D.; Cherqaoui, D. Neural Networks: Accurate non-linear QSAR Model for HEPT Derivatives. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1200-1207.
10. J. Zupan, J.; Gasteiger, J. *Neural Networks for Chemists. An Introduction*; VCH Publishers: Weinheim (Germany), 1993.
11. Mghazli, S.; Jaouad, A.; Mansour, M.; Villemin, D.; Cherqaoui, D. Neural Networks Studies: quantitative structure-activity relationships of antifungal 1-[2-(substituted phenyl)allyl]imidazoles and related compounds. *Chemosphere* **2001**, *43*, 385-390.
12. Elkhoul, K.; Afifi, A.; Kabbaj, M.; Villemin, D.; Cherqaoui, D. QSAR Analysis of Estrogen Receptor Ligands Using Neural Networks. *ACH Models in Chemistry* **2000**, *137*, 633-642.

13. Jalali-Heravi, M., Parastar, F. Use of Artificial Neural Network in a QSAR Study of Anti-HIV Activity for a Large Group of HEPT Derivatives. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 147-154.
14. Bazoui, H., Zahouily, M., Boulaajaj, S., Sebti, S., Zakarya, D, QSAR for anti-HIV activity of HEPT derivatives. *SAR QSAR Environ. Res.* **2002**, *13*, 567-577.
15. Cherqaoui, D.; Villemin, D. Use of a Neural Network to determine Boiling Point of Alkanes. *J. Chem. Soc. Faraday Trans.* **1994**, *90*, 97-102.
16. So, S.; Richards, G. Application of Neural Networks: Quantitative Structure-Activity Relationships of the Derivatives of 2,4- Diamino-5-(substituted-benzyl) pyrimidines as DHFR Inhibitors. *J. Med. Chem.* **1992**, *35*, 3201-3207.
17. Andrea, T. A.; Kalayeh, H. Application of Neural Networks in Quantitative Structure-Activity Relationships of Dihydrofolate Reductase Inhibitors. *J. Med. Chem.* **1991**, *34*, 2824-2836.
18. Cherqaoui, D.; Esseffar, M.; Villemin, D.; Cense, J. M.; Chastrette, M.; Zakarya, D. Structure-musk odour relationships studies of tetralin and indan compounds using neural networks *New J. Chem.* **1998**, *22*, 839-843.