

Exploring QSAR of non-nucleoside reverse transcriptase inhibitors by artificial neural networks: HEPT derivatives

M. Zahouily*, J. Rakik, M. Lazar, M. A. Bahlaoui, A. Rayadh, and N. Komaha

Laboratoire Catalyse, Chimiométrie et Environnement, Département de Chimie, UFR Chimie Appliquée. Université Hassan II – Mohammedia. Faculté des Sciences et Techniques de Mohammedia B.P. 146 (20650) Mohammedia, Maroc
E-mail: mzahouily@yahoo.fr

Abstract

Artificial neural networks (ANNs) can be utilized to generate predictive models of quantitative structure–activity relationships (QSAR) between a set of molecular descriptors and activity. In the present work, QSAR analysis for a set of 95 1-[(2-hydroxyethoxy)-methyl]-6-(phenylthio)thymine (HEPT) 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. The results obtained by using the NN adopted for QSAR models showing not only good statistical significance in fitting, but also high predictive ability. ($0.916 < r < 0.968$ and $q^2 = 0.8779$). The relevant factors controlling the *anti*-HIV-1 activity of HEPT 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 modelling the QSAR for HEPT derivatives

Keywords: Neural networks, QSAR, HEPT derivatives, molecular descriptor, biological activity prediction, HIV-1, reverse transcriptase

Introduction

Acquired immunodeficiency syndrome (AIDS) is caused by the depletion of helper T-lymphocytes¹ through infection by the human immunodeficiency virus² type 1 (HIV-1)³ and human immunodeficiency virus type 2 (HIV-2).⁴ Both of these retroviruses require a reverse transcriptase (RT) to convert viral RNA into proviral DNA^{5,6} that can then be inserted into the host DNA. RT has become an important target for drug discovery because of its critical role in HIV production. RT inhibitors can be divided into two major categories; nucleoside analogues and non-nucleoside inhibitors. Nucleoside analogues⁷ cause chain termination when they are incorporated within newly synthesized DNA. Non-nucleoside inhibitors block RT by binding to

a pocket adjacent to the catalytic site of the enzyme and thereby disrupt the conformation of several amino acids essential for proper RT function. In this context, the non-nucleoside reverse transcriptase inhibitors (NNRTIs)⁸ gained the greatest importance because of their specificity and their low cytotoxicity.⁹

The quantitative structure-activity relationships (QSARs) are certainly a major factor in contemporary drug design. Thus, it is quite clear why a large number of users of QSAR^{10,11} are located in industrial research units.

Artificial neural networks (ANN) are able to create internal models for complex input-output relationships based on learning from examples and therefore are useful in prediction. In protein science, ANN were successfully used to predict secondary structure^{12,13} and transmembrane segments¹⁴, the structural class^{15,16} and family^{17,18}, motifs such as co- and posttranslational modifications¹⁹, antigenic segments²⁰, signal sequence²¹ or intracellular localization.^{22,23}

ANN technique is also suited for QSAR-like applications because here a set of compounds with known activities is available for training. The essential features of ANN no linearity, adaptively, independence of statistical and other modelling assumptions, fault tolerance, universality, and real time operation make them suitable for pharmacy kinetic applications, especially where extremely complex and unfamiliar phenomena are studied.²⁴ ANN have also been applied to investigation of several QSAR.^{26,27}

The purpose of the current work is to provide an application of ANN to the structure *anti*-HIV-1 activity relationship of HEPT compounds.²⁷ The results obtained by the ANN will be compared to those given by multiple linear regressions (MLR). Thereafter, we sought to measure the contribution of each descriptor to the structure-*anti*-HIV-1 activity relationship. The NN method used here was designed to predict biological activity based on a local representation of the ligand and whole representation. The compounds of the series are represented by five properties: hydrogen-bonding acceptors (HBA), partition coefficient between *n*-octanol and water (logP), molecular weight (MW), molar refractivity (MR) and molecular energy (En).

Materials and Methods

Compounds Studied

A set of 103 HEPT compounds were taken under consideration in this study. All the molecules studied had the same parent skeleton (Figure 1).

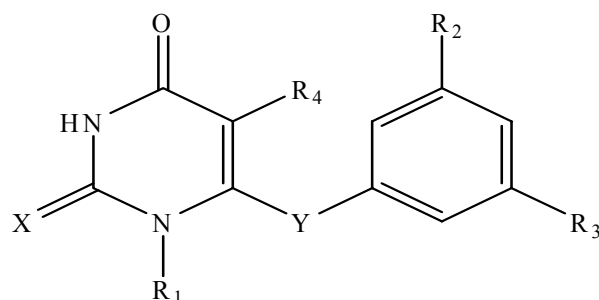


Figure 1. General structure of HEPT derivatives.

The structures and *anti*-HIV-1 activities of these compounds were described previously.^{27,28} 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 EC₅₀. The logarithm of the inverse of this parameter has been used as biological end points in the QSAR studies.

The chemical structure along with observed activity data of the compounds used in this study are shown in Table 1.

Table 1. Chemical structure of HEPT derivatives and observed anti-HIV activities

No.	R ₁	R ₂	R ₃	R ₄	X	Y	$\log\left(\frac{1}{EC_{50}}\right)_{obs.}$
1.	CH ₂ O(CH ₂) ₂ OMe	H	H	Me	O	S	5.060
2.	CH ₂ OMe	H	H	Me	O	S	5.677
3.	CH ₂ OEt	H	H	Me	O	S	6.481
4.	CH ₂ OPr	H	H	Me	O	S	5.443
5.	CH ₂ OBu	H	H	Me	O	S	5.327
6.	CH ₂ OCH ₂ Ph	H	H	Me	S	S	7.055
7.	CH ₂ OEt	H	H	Et	S	S	7.585
8.	CH ₂ OEt	Me	Me	Et	S	S	8.397
9.	CH ₂ OEt	Cl	Cl	Et	S	S	7.886
10.	CH ₂ O- <i>i</i> -Pr	H	H	Et	S	S	6.657
*11.	CH ₂ O- <i>c</i> -Hex	H	H	Et	S	S	5.795
12.	CH ₂ OCH ₂ - <i>c</i> -Hex	H	H	Et	S	S	6.455
13.	CH ₂ OCH ₂ Ph	H	H	Et	S	S	8.096
14.	CH ₂ OCH ₂ Ph	Me	Me	Et	S	S	8.154
15.	CH ₂ OCH ₂ C ₆ H ₄ (4-Me)	H	H	Et	S	S	7.107
16.	CH ₂ OCH ₂ C ₆ H ₄ (4-Cl)	H	H	Et	S	S	7.920
17.	CH ₂ OCH ₂ CH ₂ Ph	H	H	Et	S	S	7.041
18.	CH ₂ OEt	H	H	<i>i</i> -Pr	S	S	7.835
19.	CH ₂ OCH ₂ Ph	H	H	<i>i</i> -Pr	S	S	8.154

20.	CH ₂ OEt	H	H	<i>c</i> -Pr	S	S	7.022
21.	CH ₂ OEt	H	H	Et	O	S	7.721
22.	CH ₂ OEt	Me	Me	Et	O	S	8.301
23.	CH ₂ OEt	Cl	Cl	Et	O	S	8.154
24.	CH ₂ O- <i>i</i> -Pr	H	H	Et	O	S	6.468
*25.	CH ₂ O- <i>c</i> -Hex	H	H	Et	O	S	5.397
26.	CH ₂ OCH ₂ - <i>c</i> -Hex	H	H	Et	O	S	6.346
27.	CH ₂ OCH ₂ Ph	H	H	Et	O	S	8.221
28.	CH ₂ OCH ₂ Ph	Me	Me	Et	O	S	8.522
29.	CH ₂ OCH ₂ CH ₂ Ph	H	H	Et	O	S	7.017
30.	CH ₂ OCH ₂ OEt	H	H	<i>i</i> -Pr	O	S	7.920
31.	CH ₂ OCH ₂ Ph	H	H	<i>i</i> -Pr	O	S	8.522
32.	CH ₂ OEt	H	H	<i>c</i> -Pr	O	S	7.000
33.	Et	H	H	Me	O	S	5.657
34.	<i>n</i> -Bu	H	H	Me	O	S	5.920
35.	CH ₂ OCH ₂ CH ₂ OH	Me	H	Me	O	S	5.585
36.	CH ₂ OCH ₂ CH ₂ OH	Et	H	Me	O	S	5.568
37.	CH ₂ OCH ₂ CH ₂ OH	<i>t</i> -Bu	H	Me	O	S	4.920
*38.	CH ₂ OCH ₂ CH ₂ OH	CF ₃	H	Me	O	S	4.346
39.	CH ₂ OCH ₂ CH ₂ OH	F	H	Me	O	S	5.481
*40.	CH ₂ OCH ₂ CH ₂ OH	Cl	H	Me	O	S	4.886
41.	CH ₂ OCH ₂ CH ₂ OH	Br	H	Me	O	S	5.244
42.	CH ₂ OCH ₂ CH ₂ OH	I	H	Me	O	S	5.000
*43.	CH ₂ OCH ₂ CH ₂ OH	NO ₂	H	Me	O	S	4.468
*44.	CH ₂ OCH ₂ CH ₂ OH	OH	H	Me	O	S	4.086
45.	CH ₂ OCH ₂ CH ₂ OH	OMe	H	Me	O	S	4.657
46.	CH ₂ OCH ₂ CH ₂ OH	Me	Me	Me	O	S	6.585
47.	CH ₂ OCH ₂ CH ₂ OH	Cl	Cl	Me	O	S	5.886
48.	CH ₂ OCH ₂ CH ₂ OH	Me	Me	Me	S	S	6.657
49.	CH ₂ OCH ₂ CH ₂ OH	COOMe	H	Me	O	S	5.102
50.	CH ₂ OCH ₂ CH ₂ OH	COMe	H	Me	O	S	5.136
51.	CH ₂ OCH ₂ CH ₂ OH	CN	H	Me	O	S	5.000
52.	CH ₂ OCH ₂ CH ₂ OH	H	H	Et	O	S	6.958
53.	CH ₂ OCH ₂ CH ₂ OH	H	H	<i>i</i> -Pr	S	S	7.229
54.	CH ₂ OCH ₂ CH ₂ OH	Me	Me	Et	S	S	8.090
55.	CH ₂ OCH ₂ CH ₂ OH	Me	Me	<i>i</i> -Pr	S	S	8.301
56.	CH ₂ OCH ₂ CH ₂ OH	Cl	Cl	Et	S	S	7.366
57.	CH ₂ OCH ₂ CH ₂ OH	H	H	Et	O	S	6.920
58.	CH ₂ OCH ₂ CH ₂ OH	H	H	<i>i</i> -Pr	O	S	7.200
59.	CH ₂ OCH ₂ CH ₂ OH	Me	Me	Et	O	S	7.886

60.	CH ₂ OCH ₂ CH ₂ OH	Me	Me	<i>i</i> -Pr	O	S	8.522
61.	CH ₂ OCH ₂ CH ₂ OH	Cl	Cl	Et	O	S	7.853
62.	CH ₂ OCH ₂ CH ₂ OH	H	H	CH=CH-CH ₂	O	S	5.602
*63.	CH ₂ OCH ₂ CH ₂ OH	H	H	I	O	S	5.366
64.	CH ₂ OCH ₂ CH ₂ OH	H	H	CH=CH ₂	O	S	5.327
65.	CH ₂ OCH ₂ CH ₂ OH	H	H	<i>n</i> -Pr	S	S	5.000
66.	CH ₂ OCH ₂ CH ₂ OH	H	H	<i>n</i> -Pr	O	S	5.468
67.	CH ₂ OCH ₂ CH ₂ OH	H	H	H	O	S	5.154
68.	CH ₂ OCH ₂ CH ₂ OH	H	H	H	S	S	6.008
69.	CH ₂ CH=CH-Ph	H	H	Et	O	S	6.721
70.	CH ₂ CH=CH-Ph	H	H	Me	O	S	5.602
71.	CH ₂ CH=CH-Ph	Me	Me	Et	O	S	7.408
72.	CH ₂ CH=CH-thiényl	H	H	Et	O	S	7.096
73.	CH ₂ CH=CH-thiényl	Me	Me	Et	O	S	7.397
74.	CH ₂ CH=CH-furyl	H	H	Et	O	S	6.769
75.	CH ₂ CH=CH-furyl	Me	Me	Et	O	S	7.397
76.	CH ₂ CH=CH ₂ -pyridyl	H	H	Et	O	S	7.000
77.	(CH ₂) ₃ Ph	H	H	Et	O	S	6.537
78.	CH ₂ OCH ₂ CH ₂ OH	H	H	Et	O	CH ₂	6.455
79.	CH ₂ OCH ₂ CH ₂ OH	Me	Me	Et	O	CH ₂	7.886
80.	CH ₂ OEt	H	H	Et	O	CH ₂	7.387
81.	CH ₂ OEt	Me	Me	Et	O	CH ₂	8.795
82.	CH ₂ OCH ₂ CH ₂ OH	H	H	<i>i</i> -Pr	O	CH ₂	7.200
83.	CH ₂ OCH ₂ CH ₂ OH	Me	Me	<i>i</i> -Pr	O	CH ₂	8.568
84.	CH ₂ OEt	H	H	<i>i</i> -Pr	O	CH ₂	7.376
85.	CH ₂ OEt	Me	Me	<i>i</i> -Pr	O	CH ₂	9.221
86.	<i>n</i> -Bu	H	H	Et	O	CH ₂	6.677
87.	<i>n</i> -Bu	H	H	<i>i</i> -Pr	O	CH ₂	7.376
88.	CH ₂ CH ₂ OMe	H	H	Et	O	CH ₂	6.602
89.	CH ₂ CH ₂ OMe	H	H	<i>i</i> -Pr	O	CH ₂	7.284
90.	CH ₂ OCH ₂ OMe	H	H	Me	O	CH ₂	4.638
91.	CH ₂ OCH ₂ CH ₂ OH	<i>c</i> -Hex	H	Me	O	S	5.455
92.	CH ₂ OCH ₂ CH ₂ OH	H	H	Me	O	O	5.050
93.	CH ₂ OCH ₂ CH ₂ OH	H	H	Me	O	S	5.244
*94.	CH ₂ SCH ₃	H	H	Et	O	CH ₂	8.689
95.	CH ₂ SCH ₂ CH ₃	H	H	Et	O	CH ₂	7.398
96.	CH ₂ SCH ₃	Me	Me	Et	O	CH ₂	7.301
97.	CH ₂ SCH ₂ CH ₃	Me	Me	Et	O	CH ₂	8.398
98.	CH ₂ SCH ₃	H	H	<i>i</i> -Pr	O	CH ₂	7.699
99.	CH ₂ SCH ₂ CH ₃	H	H	<i>i</i> -Pr	O	CH ₂	8.222

100.	CH ₂ OCH ₂ CH ₂ Cl	H	H	Me	O	S	5.820
101.	CH ₂ OCH ₂ CH ₂ N ₃	H	H	Me	O	S	5.240
102.	CH ₂ OCH ₂ CH ₂ F ₃	H	H	Me	O	S	5.690
103.	CH ₂ OCH ₂ CH ₂ OCOPh	H	H	Me	O	S	5.120

log (1/EC₅₀) observed values were used as dependant variable in which EC₅₀ represents the molar concentration of drug required to achieve 50% protection of MT-4 cells against effect of HIV-1. The molecules designated with an asterisk (*) were removed later based on their di value (see text below) to perform a new ANN model derived from the remains 95 subset data set.

Molecular Descriptors Used

Each molecule was described by 4 local descriptors which are: parameters measuring Molar refractivity (MR), Molecular weight (MW), Hydrophobic (water/octanol partition coefficient: log P), Hydrogen-bonding acceptors (HBA) and one whole descriptor: the molecular energy (En).

Neural Network

All the feed-forward NN used in this paper are three-layer networks with the first (input) layer having five neurones, representing the relevant descriptors obtained in MLR technique.²⁹ Although there are neither theoretical nor empirical rules to determinate the number of hidden layers or the number of neurone layers, one hidden layer seems to be sufficient in the most chemical application of ANN. The output layer represents the calculated activity values log (1/EC₅₀). Figure 2a show a typical architecture of such ANN. This is formed in three layers, called the input layer, hidden layer, and output layer. Each layer consists of one or more neurons, represented in this diagram by the small circles. The line between the neurons indicates the flow of information from one neuron to the next. The neurons of the input layer are passive while those of the hidden and output layers are active. The variables: X₁₁, X₁₂ ..., X₁₅ (see text below) hold the data to be evaluated. Each value from the input layer is duplicated and sent to all of the hidden neurons (fully interconnected structure). As shown in Fig. 2b, the values entering the hidden neurons are multiplied by weights (the W_i values). The weighted inputs are then added to produce a single number (shown in the diagram by the symbol Σ). Before leaving the neuron, this number is passed through a nonlinear mathematical function called sigmoid. This is an “S” shaped curve that limit the neuron’s output between 0 and 1.

Results and Discussion

Computation

In a back-propagation ANN 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. It seems that, for most applications of the ANN to chemistry, one hidden layer is sufficient.³⁰ However, for the determination of the number of hidden neurons, some authors^{31,32} have proposed a parameter ρ , which plays a major role in determining the best ANN architecture. $\rho = (\text{Number of data points in the training set} / \text{Sum of the number of connection in the NN})$.

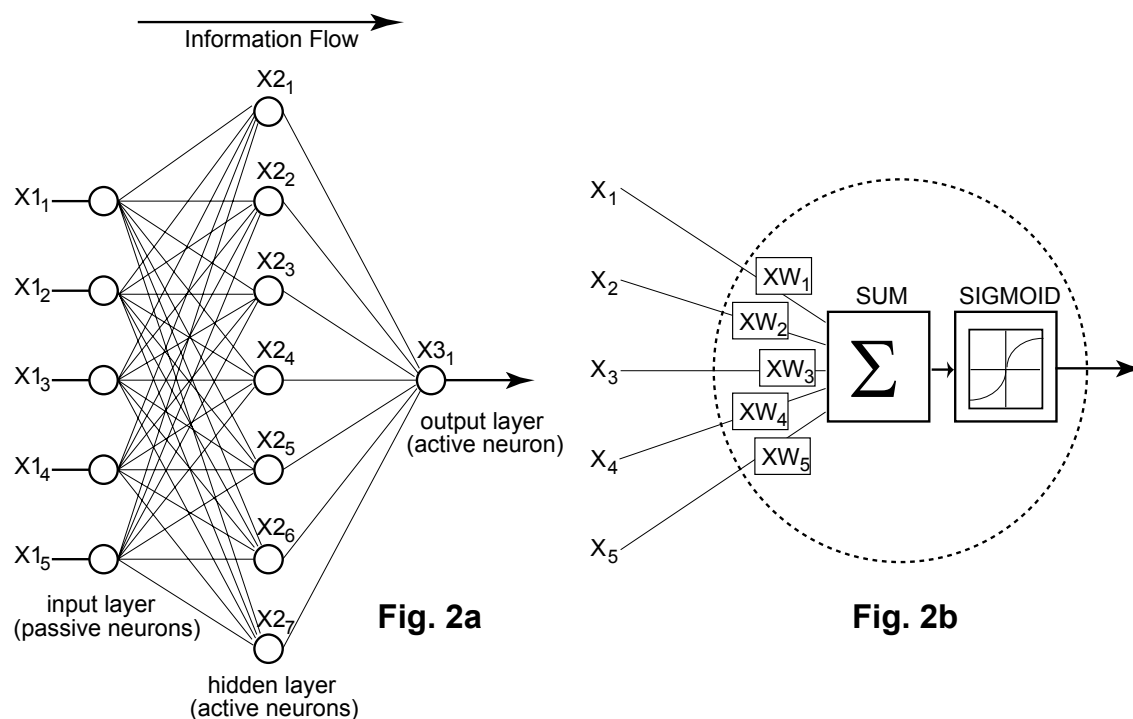


Figure 2. Typical architecture of a three layer artificial neural network (a) and flow diagram of the active neurons used in the hidden and output layers (b).

MLR was performed on all compounds described in Table 1, a few suitable models were obtained and the pertinent descriptors appearing in the best one were selected to perform the ANN model (see Table 2).

To remain close to the experimental error (5%), we also take away the 8 molecules having $d_i = \frac{(|obs_i - cal_i|)}{obs_i}$ higher than $\frac{2s}{(\overline{\log 1/EC_{50}})}$ where d_i is the value of descriptor i and $\overline{(\log 1/EC_{50})}$ is the mean of observed activity. The molecules that were removed are labelled with an asterisk in Table 1. Consequently, a new model was derived on 95 molecules. The [5-7-1] neural network architecture was developed with the optimum momentum 0.9 and with 10000 iterations. The seven hidden neurone were chosen to maintain ρ between 1 and 3 ($1 < \rho < 3$). To verify this condition we have also performed a trial by taking five to thirteen neurons in the hidden layers. The best results for the training and tests are given in the Table 3. The best model

is that corresponding to the optimum r and SEC parameters. A hidden layer of seven neurones was selected.

Table 2. Some of significant MLR models obtained for all compounds described in Table 1 and percentage of $\log(1/EC_{50})$ variable's variance explained by the models (r^2)

No. ^a	Descriptors with significant regression coefficient from higher to lowest ^b	s ^c	r ²
1	+ Log P(R ₃)***; - MW(R ₃)**; + MR(R ₄)**; - HBA(R ₄)*; + En	0.54	0.899
2			
3	- MR(R ₄)***; - MW(R ₃)**; + log P(R ₃)**; - HBA(R ₄)*; + En		0.975
4			
5			

^a The model N° 3 was selected to perform the ANN model

^b Only the sign of the regression coefficient is shown. MR = Molar refractivity; MW = Molecular weight; log P = Hydrophobic water/octanol partition coefficient; HBA = Hydrogen-bonding acceptors; En = Molecular energy

^c Standard deviation

* $p \leq 0.05$; ** $p \leq 0.01$; *** $p \leq 0.001$

Table 3. Standard error of computation (SEC) and correlation coefficient obtained by NN trained with 95 data points

Architecture	r	SEC
5-5-1	0.9662	0.4271
5-6-1	0.9667	0.4238
5-7-1	0.9684	0.4135
5-8-1	0.9655	0.4310
5-9-1	0.9657	0.4297
5-10-1	0.9624	0.4504
5-11-1	0.9679	0.4166
5-12-1	0.9655	0.4311
5-13-1	0.9657	0.4299
MLR	0.9163	0.4928

With the above architecture, a correlation coefficient of 0.968 ($n = 95$) between calculated and observed $\log(1/EC_{50})$ with a standard error of 0.413 was obtained. In comparison (see also figure 3), the model established with MLR gave a correlation coefficient of 0.916 and a standard error of 0.493.

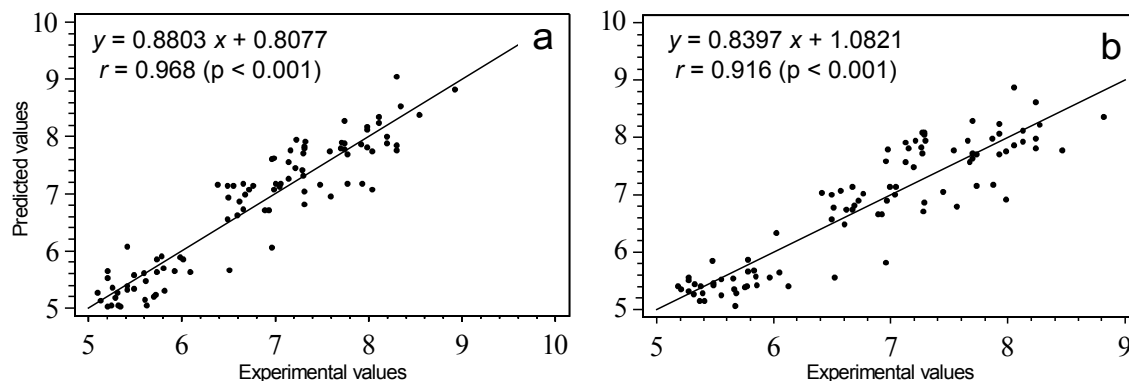


Figure 3. Observed versus calculated *anti*-HIV activities obtained with the ANN (a) and MLR (b) models.

Based on this result, a comparison of the quality of the MLR and ANN models shows that the ANN models have substantially better predictive capability because the ANN approach gives better results than MLR. The standard errors of calculation are lower and the correlation coefficients are higher with ANN 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.

Assessment of the neural network validity and predictive ability

The ability of the network to learn the data and predict the biological activity was tested by the “leave-one-out” cross-validation.³³ In this procedure one compound is removed from the 95 subsets data set. And the output of the removed compound was predicted for each subset. The cross-validation coefficient q^2 was then calculated according to the following equation.³⁴

$$q^2 = 1 - \left(\frac{PRESS}{Variance} \right)$$

where PRESS is the predictive residual sum of squares. They yielded a $q^2 = 0.8779$ indicating a good predictive quality of the model, according to Wold.³⁵

Analysis of descriptor's contribution in ANN model

The contribution of descriptors i ($i = 1-5$) was estimated from the trained [5-7-1] configuration network. The descriptor under study was removed from the [5-7-1] trained ANN together with its corresponding weights. Then the network [4-7-1] calculated the output of each molecule as usual. The mean of the deviations absolute values Δm_i between the observed activity and the estimated one for all compounds was calculated. This process was reiterated for each descriptor. Finally, the contribution C_i ³⁶ of descriptor i is given by:

$$C_i = 100 \frac{\Delta m_i}{\sum_{i=1}^4 \Delta m_i}$$

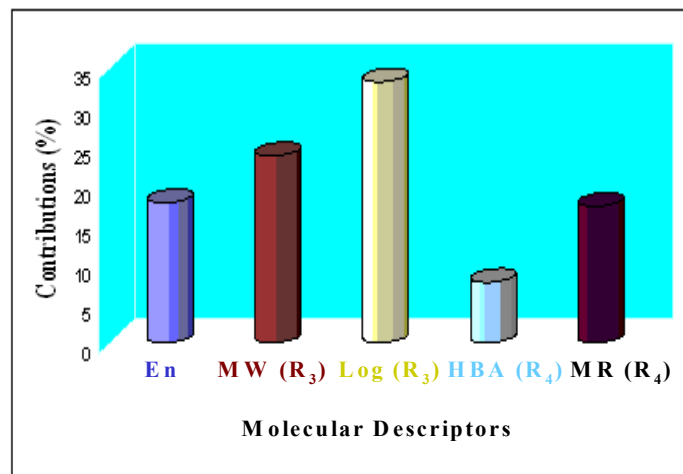


Figure 4. Descriptor's contribution in the ANN model.

Figure 4 indicate that the relative importance of the descriptors varied in the following order: $\log P(R_3) > MW(R_3) > En > MR(R_4) > HBA(R_4)$.

The comparison of the quality of the MLR and ANN models (Figure 5) shows that the latter have substantially better predictive capabilities.

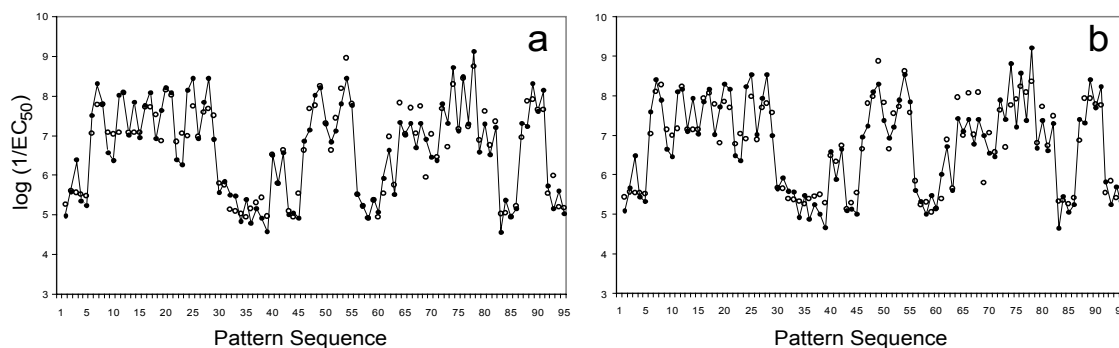


Figure 5. Observed (closed circles) and predicted (open circles) activity values with the ANN (a) and MLR (b) models.

Conclusion

The results obtained by ANN method for QSAR indicate that ANN was able to establish a satisfactory relationship between the molecular descriptors and the *anti*-HIV activity. The main factor controlling the *anti*-HIV activity of HEPT derivatives inhibiting HIV-1 reverse transcriptase have been determined by ANN.

References

1. Alimonti, J.B.; Ball, T.B.; Fowke, K.R. *Journal of general Virology* **2003**, 84, 1649.
2. Gottlieb, M.S.; Schroff, R.; Schanker, H.M. *New England Journal of Medicine* **1981**, 305, 1425.
3. Arré-Sinoussi, F.; Chermann, J.C.; Rey, F.; Nugeyre, M.T.; Chamaret, S.; Gruest, J.; Dauguet, C.; Axler-Blin, C.; Brun-Vezinet, F.; Rozioux, C.; Rozenbaum, W.; Montagnier, L. *Science* **1983**, 220, 869.
4. Clavel, F.D.; Guetard, D.; Brun-Vezinet, F.; Chamaret, S.; Rey, F.; Santos-Ferreira, M.O.; Laurent, A.G.; Danguet, C.; Katlama, C.; Rosioux, C. *Science* **1986**, 233,343.
5. Le Grice, S.F.J.; A.M.S.S.P.G. Eds, Cold Spring Harbor laboratory press, New York **1993**, pp 163-191.
6. Basu, A.; Basu, S.; Modak, M.J. *Biochemical and Biophysical Research Communications* **1992**, 183(3), 1131.
7. Mitsuya, H.; Broder, S. *Proc. Natl. Acad. Sci. U. S. A.* **1986**, 83, 1911.
8. De Clercq, E. *Antiviral Res.* **1998**, 38, 153.
9. De Clercq, E. *Il Farmaco* **1999**, 54, 26-45.
10. Trinajstic, N. *Chemical Graph Theory*, 2nd Edn.; Boca Raton, F.L. Eds.; CRC Press, **1992**; pp 20.
11. a) Bazoui, H.; Zahouily, M.; Boulaajaj, S.; Sebti, S.; Zakarya, D. *SAR QSAR Environ. Res.* **2002**, 13, 567, b) Bazoui, H.; Zahouily, M.; Sebti, S.; Boulaajaj, S.; Zakarya, D.J. *Mol. Model.* 8(2002)1-7, c) Agrawala, V. K.; Singha, J.; Mishraa, K. C.; Khadikar, P. V.; Jaliwalac Y. A. *Arkivoc* **2006**, ii, 162.
12. Rost, B.; Sander, C. *Proc. Natl. Acad. Sci. USA.* **1993**, 90, 7558.
13. Kaur, H.; Raghava, G.P.S. *Protein Science* **2003**, 12, 627.
14. Lohmann, R.; Schneider, G.; Behrens, D.; Wrede, P. *Protein Science* **1994**, 3, 1597.
15. Bohr, J.; Bohr, H.; Brunak, S.; Cotterill, R.M.J.; Fredholm, H.; Lautrup, B.; Petersen, S.B. *J. Mol. Biol.* **1993**, 231, 861.
16. Dubchak, I.; Muchnik, I.; Holbrook, S.R.; Kim, S.H. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, 92, 8700.
17. Wu, C. H. *Methods Enzymol.* **1996**, 266, 71.
18. Wu, C.H.; Zhao, S.; Chen, H.L.; Lo, C.J.; McLarty, J. *Search. Comput. Appl. Biosci.* **1996**, 12, 109.

19. Julenius, K.; Molgaard, A.; Gupta, R.; Brunak, S. *Glycobiology* **2004**, 15(2), 153.
20. Nielsen, M.; Lundegaard, C.; Worning, P.; Lauemoller, S.L.; Lamberth, K.; Buus, S.; Brunak, S.; Lund, O. *Protein Science* **2003**, 12,1007.
21. Jagla, B.; Schuchhardt, J. *Bioinformatics* **2000**, 16(3), 245.
22. Schneider, G.; Wrede, P. *J. Mol. Evol.* **1993**, 36, 586.
23. Von Heijne, G. In: Concepts in Protein Engineering and Design; Wrede, P.; Schneider, G. Eds.; Berlin, New York, **1994**; pp. 263-279.
24. Erb, R.J. *Pharm. Res.* **1993**, 10, 165.
25. Zahouily, M.; Rhihil, A.; Bazoui, H.; Sebti, S.; Zakarya, D. *J. Mol. Model.* **2002**, 8, 168.
26. Zahouily, M.; Rayadh, A.; Aadil, M.; Zakarya, D. *J. Mol. Model.* **2003**, 9,242.
27. Miyasaka, T.; Tanaka, H.; Baba, M.; Hayakawa, H.; Walker, R.T.; Balzarini, J.; De Clercq, E. *Journal of Medicinal Chemistry*, **1989**, 32, 2507.
28. Tiwari, B. K.; Thakur, A.; Thakur, M.; Pandey, N. D.; Narvi, S. S.; Thakur, S. *Arkivoc* **2006**, ii, 213.
29. Garg, R.; Gupta, S. P.; Gao, H.; Babu, M. S.; Debnath, A. K. *Chem. Rev.* **1999**, 99, 3525.
30. Zupan, J.; Gasteiger, J. *Neural Networks for Chemists. An Introduction*; VCH: Weinheim (Germany), **1993**.
31. So, S.; Richards, G. *J. Med. Chem.* **1992**, 35, 3207.
32. Andrea, T. A.; Kalayeh, H. *J. Med. Chem.* **1991**, 34, 2824.
33. Rumhelart, D.E.; Hinton, C.E.; Williams, R.J. *Nature* **1986**, 406, 323, 533.
34. Tetko, I.V.; Villa, A.E.P.; Livingstone, D.J. *J. Chem. Inf. Comput.* **1996**, 393, 36, 794.
35. Wold, S. *Quant. Struct.-Act. Relat.* **1994**, 10, 191.
36. Cherquaoui, D.; Esseffar, M.; Villemin, D.; Cence, J.M.; Chastrette, M.; Zakarya, D. *New J. Chem.* **1998**, 839.