

A QSAR study investigating the potential anti-HIV-1 effect of some Acyclovir and Ganciclovir analogs

Anna Maria Almerico,* Marco Tutone, and Antonino Lauria

*Dipartimento Farmacochimico Tossicologico e Biologico, Università di Palermo,
Via Archirafi 32, 90123 Palermo
E-mail: almerico@unipa.it*

Dedicated to Professor Nicolò Vivona on the occasion of his 75th birthday

Abstract

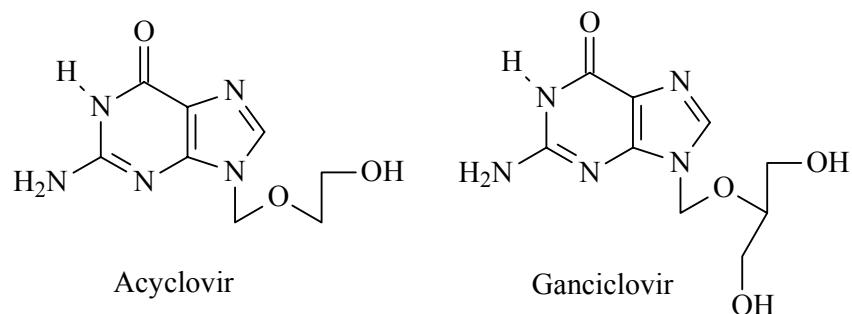
A QSAR study, involving the use of calculated physical-chemical properties (TSARTM), and the use of a neural network approach (TSARTM), has been performed on the potential anti-HIV-1 activity of a series of Acyclovir and Ganciclovir analogs. Model obtained allows reliable predictions for the anti-HIV-1 activity of these derivatives, and showed that the presence of the Ganciclovir chain in triazolopyrrolopyrimidine and pyrimidopyrrolopyrimidine series seems to increase the antiviral effect.

Keywords: QSAR, neural network, BMLR, anti-HIV-1 activity

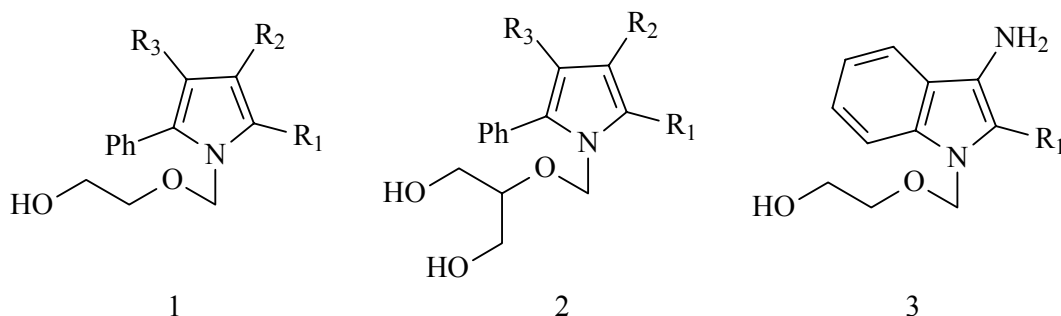
Introduction

Nucleoside analogs, such as AZT, Lamivudine, Zalcitabine, Didanosine, are again important drugs for the treatment regimen of HIV infections. In fact, most 3- and 4-drug regimens involve 2 nucleoside analogs, chosen on the basis of convenience, potential side effects, and patient preference.

Structural modifications of these classes of compounds involved both glycoside and aglycone moieties. As a result of these studies, by mean of simplification of riboside portion, Acyclovir and Ganciclovir, acyclic glycosides related to 2'-Deoxyguanosine, could be obtained (Figure 1). Acyclovir (ACV) is highly effective against herpes simplex virus (HSV) and varicella zoster virus (VZV), while Ganciclovir, an analog of Acyclovir, is about as potent as ACV against these viruses, but showed higher activity against human cytomegalovirus (HCMV), an important pathogen in immunocompromised and acquired immune deficiency syndrome patients. These compounds also showed weak activity as anti-HIV-1 agents, and in our previous studies on HIV-1 inhibitors, we synthesized some glycosidoheterocycles.¹⁻⁴

**Figure 1**

These glycosidopyrroles of type **1** or **2** and glycosidoindoles **3** (Fig. 2), being related to Acyclovir and Ganciclovir because of the acyclic moiety, could be considered more simple carbon bioisosters of the above mentioned antiviral drugs.

**Figure 2**

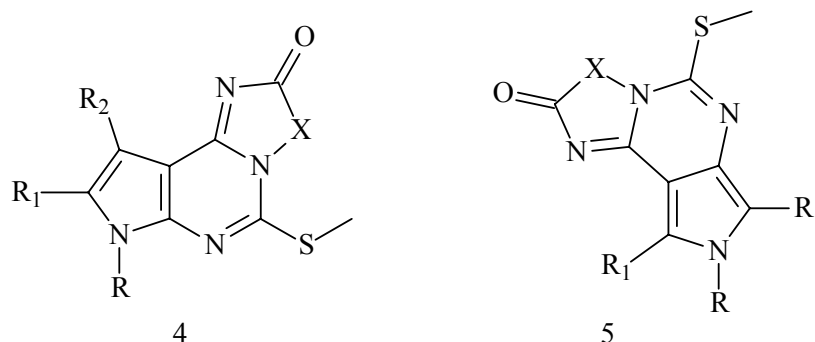
In this series, the 1-hydroxyethoxymethylpyrrole **1a** showed weak anti-HIV-1 activity and resulted cytotoxic, whereas only the 3-nitro-2,4,5-triphenylpyrrole derivative **1i** inhibited HIV-1 replication at concentration non-cytotoxic for MT-4 cells. Acyclic glycosidopyrroles of type **2** resulted less cytotoxic than derivatives of type **1**, but also less selective against HIV-1. No significant activity could be evidenced for most of the benzofused derivatives of type **3**.

Table 1

	R ₁	R ₂	R ₃	Log 1/EC ₅₀ ^a		R ₁	R ₂	R ₃	Log 1/EC ₅₀ ^a
1a	Me	COOEt	Cl	4.954	2c	Ph	CN	Br	4.699
1b	Me	COOEt	NO ₂	4.477	2d	Ph	CN	NO ₂	4.699
1c	Ph	H	NO ₂	5.342	2e	Ph	COOEt	Br	4.903
1d	Ph	Br	Br	5.279	2f	Ph	COOEt	NO ₂	4.778
1e	Ph	COOEt	Br	5.204	2g	Me	COOEt	NO ₂	4.699
1f	Ph	COOEt	NO ₂	4.778	2h	Ph	CN	NH ₂	4.699
1g	Ph	CN	NO ₂	4.477	2i	Ph	Ph	NH ₂	5.819
1h	Ph	Ph	Br	5.819	2j	Ph	COOEt	NH ₂	4.699
1i	Ph	Ph	NO ₂	5.431	2k	Me	COOEt	NH ₂	4.699
1j	Ph	COOEt	NH ₂	NA	2l	Ph	Ph	N ₃	5.301
1k	Ph	CN	NH ₂	4.477	2m	Ph	CN	N ₃	4.699
1l	Ph	Ph	NH ₂	5.380	2n	Ph	COOEt	N ₃	4.699
1m	Ph	COOH	NO ₂	4.477	2o	Me	COOEt	N ₃	4.699
1n	Ph	PH	N ₃	5.176	3h	Ph			4.954
2a	Ph	Ph	Br	5.518	3i	2-Pyr			5.301
2b	Ph	Ph	NO ₂	5.342					
2c	Ph	CN	Br	4.699	Acyclovir				5.000
2d	Ph	CN	NO ₂	4.699	Ganciclovir				5.301

^aEC₅₀ is expressed in μM; NA, not available

In an attempt to optimize the future synthetic work in the series of pyrrole containing heterocycles, these glycosido derivatives, for which anti-HIV-1 activity data are available (Table 1), were subjected to a QSAR (Quantitative Structure-Activity Relationships) analysis, in order to determine the physico-chemical properties which appear to influence biological activity and to predict potential activity of derivatives of the new ring systems of type **4** and **5** (Figure 3), analogs of Acyclovir and Ganciclovir.



X = NMe, CH₂, CHMe, CH₂CH₂, CH(Me)CH₂; R = H, CH₂O(CH₂)₂OH, CH₂OCH(CH₂OH)₂, R₁ = H, Me, Ph, R₂ = H, Me, Ph.

Figure 3

Results and Discussion

The activity data, expressed as $\log 1/EC_{50}$ values [where EC_{50} is the compound concentration (μM) required to reduce the virus-induced cytopathogenicity (HIV-1) by 50%], were used as the dependent variable in the QSAR study. All 3D-structures were generated and minimized using semi-empirical quantum mechanics, utilizing AM1 Hamiltonian within the TSAR package.⁵ For each whole molecule more than 60 descriptors (electronic, shape, thermodynamic, and topological) were generated. Correlation matrix of these descriptors revealed correlation coefficient > 0.9 for most of them, so to avoid redundant information in further analysis, only thirteen descriptors were selected: accessible surface area (ASA); heat of formation (ΔH); E_{LUMO} ; E_{HOMO} ; total dipole (μ); ellipsoidal volume; $\log P$; total lipole; Kier $^1\chi^v$ index; shape flexibility index (φ); Balaban index; number of H-bond donors (HBD); number of H-bond acceptors (HBA). The selected set of molecular descriptors identifies the key molecular characteristics that better give account for the biological activity.

A best multilinear regression (BMLR) model was carried out and the final equation includes six out of the thirteen descriptors (eq. 1). The statistic for this equation are $N=32$, $n=6$, $R^2=0.782$, $R^2_{\text{cvLOO}}=0.667$, $F=14.92$, $s=0.2$, $\text{RandTest} = 0.251$ (Figure 4). Although R^2_{cvLOO} is valuable, not all the selected descriptors showed a linear relationship with the activity data, therefore because of a non-linear behavior they were excluded.

$$\text{Log } 1/EC_{50} = -3.0769(D1) + 0.0692(D2) - 0.0028(D3) - 0.3632(D4) - 0.2418(D5) + 0.1098(D6) + 10.4 \quad \text{eq. 1}$$

Where:

D1 = Balaban index, standard error = 0.62

D4 = HBD, standard error = 0.12

D2 = Total lipole; standard error = 0.014

D5 = $\log P$, standard error = 0.10

D3 = ΔH , standard error = 0.0008

D6 = φ , standard error = 0.093

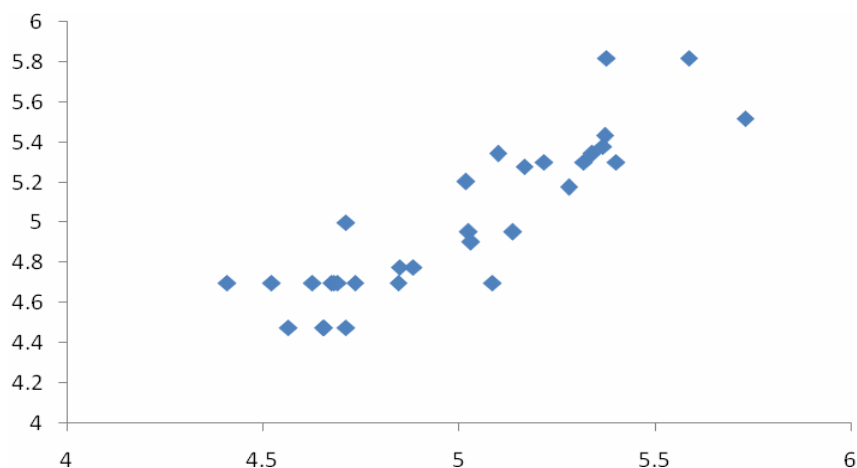


Figure 4. Predicted $\log 1/EC_{50}$ (x axes) vs Experimental $\log 1/EC_{50}$ (y axes) for BMLR model

In an attempt to improve predictions obtained for the HIV-1 inhibitory data in the above QSAR model and to find relationships among all the descriptors of our interest, a neural net analysis has been performed using the functionality offered by TSAR software packages. The multiple-layer forward feed neural network (FFNN) functionality, which undergoes a supervised training by the back propagation error, was used.⁶ In our case, the inputs for the neural network were the descriptors obtained above, while the outputs were the $\log 1/EC_{50}$ values. The FFNN functionality within the TSAR software automatically computes the number of the hidden neurons, as well as the number of training and test patterns (in this case 70% was used for training, 30% for test purposes). The number of neurons in the hidden layer and the number of rows in the training set are balanced to achieve the optimum predictive power for the neural network. The statistic obtained for the FFNN treatment of the HIV-1 data were $N = 32$, input columns (descriptors) = 13, net configuration = 13-2-1 (13 input nodes, 2 processing nodes, 1 output node), with Test root mean square = 0.145 and $R^2 = 0.89$ (Figure 5).

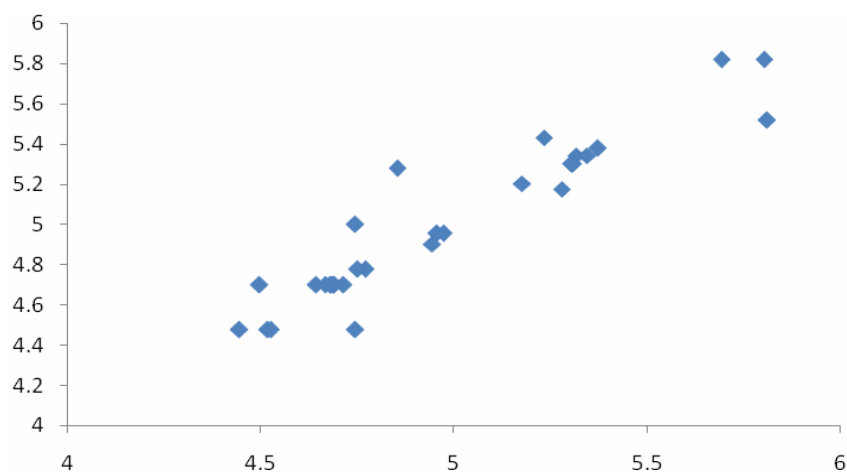


Figure 5. Predicted $\log 1/EC_{50}$ (x axes) vs Experimental $\log 1/EC_{50}$ (y axes) for FFNN model

The QSAR model obtained exhibits strong dependencies on the directional components to lipophilicity (total lipole), such as dipole and E_{HOMO} , while topological index, HBA and ΔH showed a decreasing relationship with the output values (Figure 6).

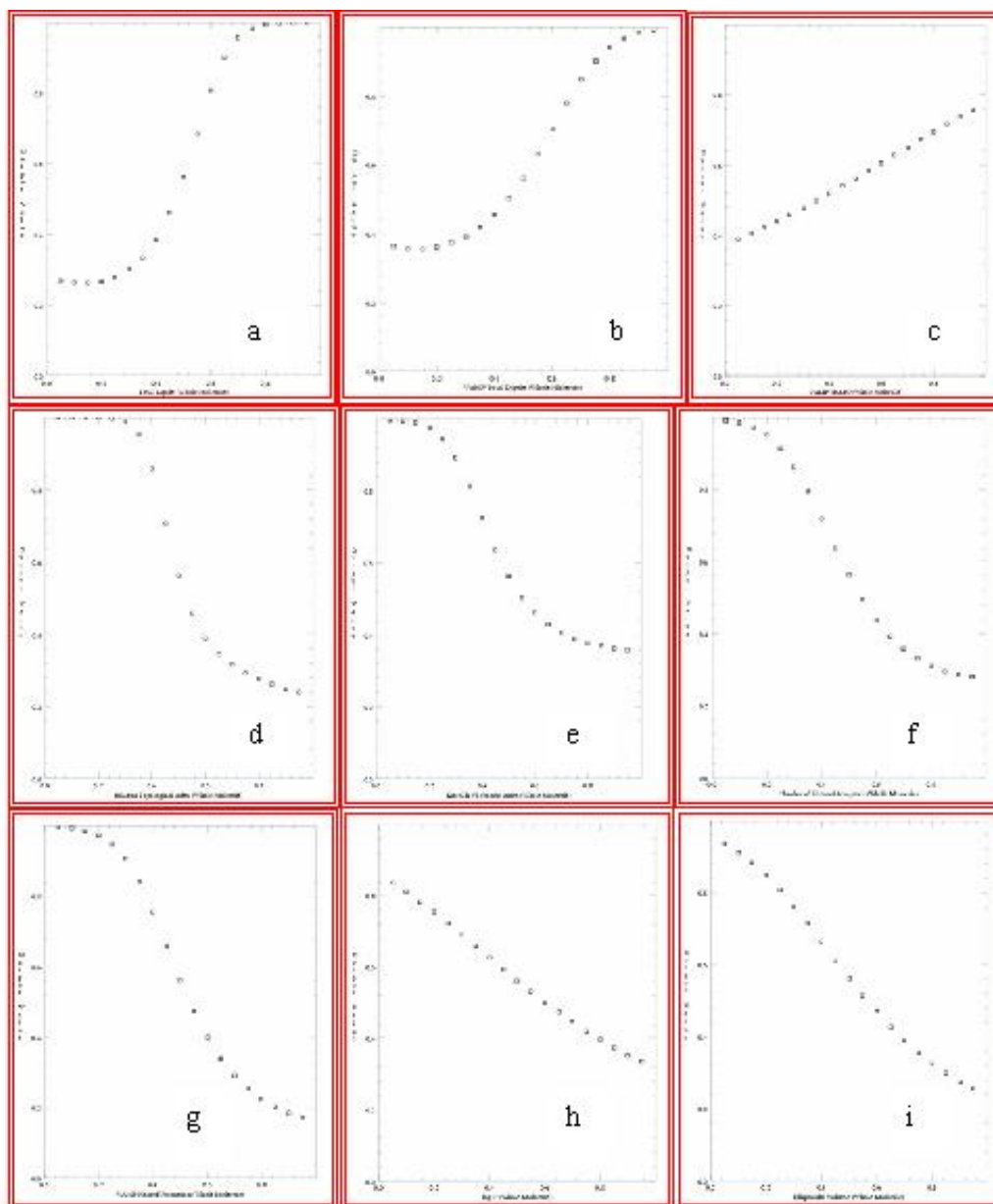


Figure 6. $\log 1/EC_{50}$ (y axes) vs: (a) total lipole, (b) μ , (c) E_{HOMO} , (d) Balaban index, (e) Kier $1\chi^v$, (f) H-bond acceptors, (g) ΔH , (h) $\log P$, (i) ellipsoidal volume (x axes)

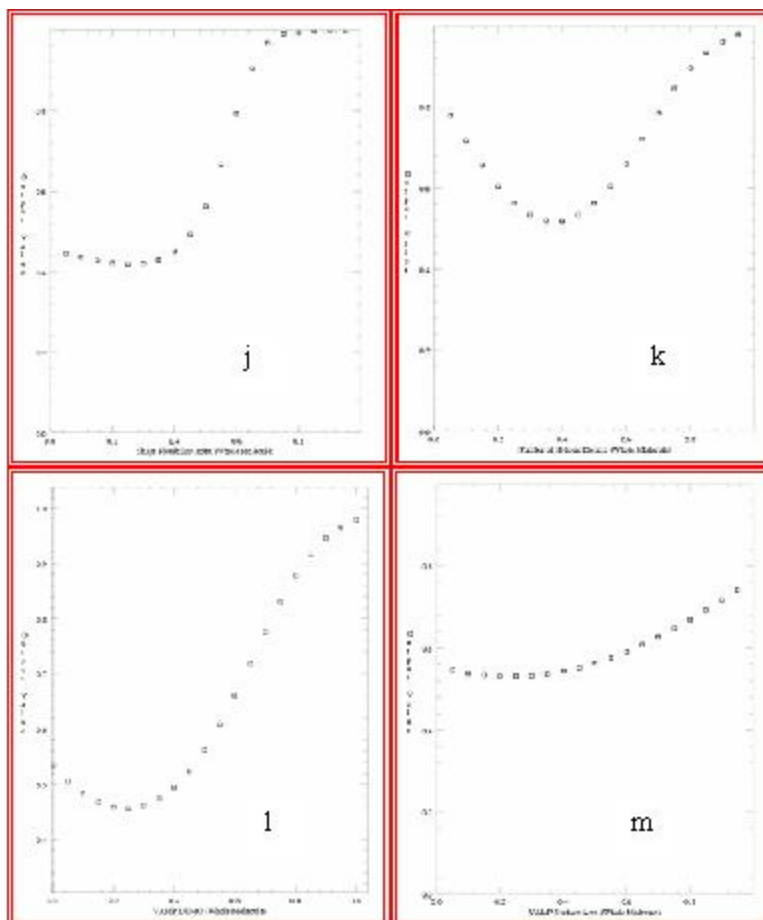


Figure 6. (Continued) log 1/EC₅₀ (y axes) vs: j) ϕ , k) H-bond donors, l) E_{LUMO}, m) ASA (x axes)

Log P and Ellipsoidal volume also showed a decreasing relationship with the output values, but in a quasi-linear fashion. For the other descriptors, it resulted impossible to find a simple relationship (Figure 6).

However although using the same descriptors for the BMLR model, the FFNN treatment appears to improve the predictions obtained. All the considered descriptors were included into the second analysis, in fact the advantage of such a statistical treatment is that all non-linear relationships must be incorporated explicitly into a regression model, in contrast, FFNN make no assumptions about a linear dependence of the input variable. Non-linear terms are built into the net as a function of the network topology. Therefore the FFNN model could be used to obtain a realistic prediction of activity data for derivatives of classes **4** and **5** (Figure 3).

The results of this analysis evidenced that several derivatives of type **4** are expected to show a higher activity than compounds **1**, **2** and **3**. In particular, in the triazolopyrrolopyrimidine series (X = NMe) and in the pyrimidopyrrolopyrimidine one [X = CH₂CH₂, CH(Me)CH₂], the presence of the Ganciclovir chain seems to be relevant for the appearance of biological activity.

The structures of the more interesting compound, for which the predicted activity was found in the range $\text{Log } 1/\text{EC}_{50} = 5.814\text{-}5.817 \mu\text{M}$, are shown in Figure 7. Further the aryl moieties in position 4 and 5 of the pyrrole ring in this type of derivatives mirror the experimental activity data obtained for compounds of series **1** and **2**, where the most active derivatives present the Ganciclovir chain and at least two aromatic substituents.

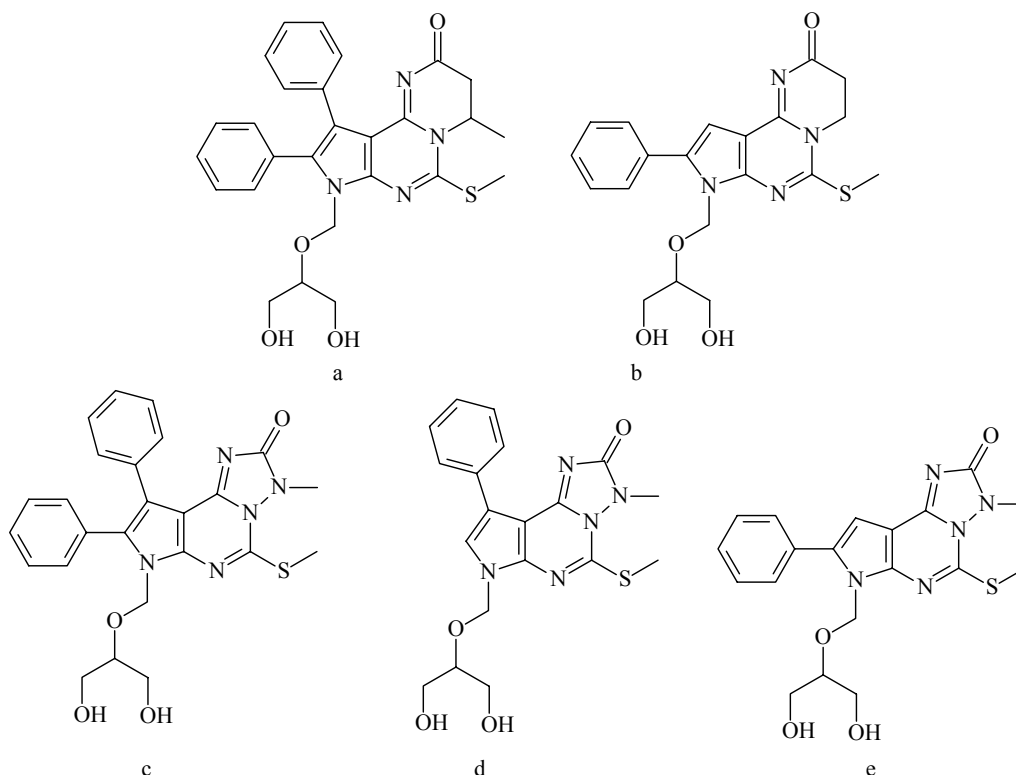


Figure 7

Conclusions

The result of our QSAR study allows to identify physico-chemical descriptors which can be strictly related to the anti-HIV-1 activity of these Acyclovir and Ganciclovir analogs. The improved predictive ability of FFNN model respect the BMLR makes the neural network the model of choice to obtain a realistic prediction of activity for the new derivatives of type **4** and **5**, which are currently under investigation.

Experimental Section

Molecular descriptors. The compounds were sketched as 2D representations using ChemWindow and the 3D structures were optimized by semiempirical methods (AM1 Hamiltonian) using CORINA and COSMIC modules from TSAR 3.2 software.⁷

A selected set of molecular descriptors, which identify the molecular characteristics that can be related to the biological activity, was calculated. The selected thirteen descriptors were: VAMP accessible surface area (ASA); VAMP heat of formation (ΔH); VAMP E_{LUMO} ; VAMP E_{HOMO} ; total dipole (μ); ellipsoidal volume; log P; total lipole; Kier $^1\chi^v$ index; shape flexibility index (ϕ); Balaban index; number of H-bond donors (HBD); number of H-bond acceptors (HBA). In particular ellipsoidal volume, which is defined by the moments of inertia, and accessible surface area give information about steric properties of a molecule, log P is a typical QSAR variable, related to hydrophobic-hydrophilic profile of the inhibitor, total lipole is a measure of the lipophilic distribution in a 3D space and it is calculated from the summed atomic log P values, the number of H-bond donors and acceptors give other information about the ability of the inhibitor to stabilize its interaction with the receptor active site, total dipole gives information about the electronic features, E_{LUMO} and E_{HOMO} are energetic variables that classify the set of the inhibitors in terms of their ability to act as electrophiles and nucleophiles. Other energy variables are heat of formation, which classifies the inhibitors in terms of relative thermodynamic stability and is widely used in chemometric studies, and accessible surface area in water, which classifies the behaviour of the derivatives in the physiological solvent. The set of topological indexes, such as flexibility index Φ , is related with the degree of linearity and the presence of cycles and/or branching; whereas Balaban index, and Kier-Hall topological index $^1\chi^v$ encoded information about the flexibility, size, branching and shape of the molecules.

Forward feed neural network (FFNN). In this type of regression analysis a net is trained to predict dependent variables from a set of explanatory variables. The architecture of a neural net is highly flexible in terms of the number of input and output layers, the number of neurons within each layer, the extent of connections within the network and the directional transfer of information. In our case the first layer in the net comprises input nodes that receive input data and transfer this to the first step of processing neurons. A computer neuron performs the same basic function as a biological neurons, i.e. a summation of the combined inputs that is mapped to an output value via a transfer function. The neural net implementation in TSAR uses an identity activation function.⁸ Initially, a weighting value of 1.0 is applied to all connections. Then a Monte Carlo algorithm is run to select a better set of starting weights within the constrain limits. The multiple-layer forward feed net undergoes supervised training using the method of back propagation: input values are propagated in a forward direction through the net such that an output is calculated for each node based on the current weightings. For the given set of data (complete with tested input and output data) the net calculates the difference between this output value and the experimental output. This error is then used to adjust the weights of the previous

layers. Progressively (*via* several thousand iterations), the weights within the neural net are adjusted to minimize the overall error. The “knowledge” of a trained net is a function of the strengths of the connections within the network. A proportion of the input data will be excluded from the training set and used as a test set. The output values for these data are predicted simultaneously as the training progresses and the error reported. This procedure allows to assessing the predictive power of the training network and avoids difficulties with overtraining. The convergence of the data during the calculation was monitored and the dependencies of the output variable on each of the input variable, as a series of dependence plots, were displayed and analyzed.

Supplementary Information Available

Tables of selected descriptors and of calculated EC₅₀ values for the derivatives included in this study are available as supporting information.

References and Notes

1. Diana, P.; Barraja, P.; Almerico, A. M.; Dattolo, G.; Mingoia, F.; Loi, A. G.; Congeddu, E.; Musiu, C.; Putzolu, M.; La Colla, P. *Il Farmaco* **1997**, *52*, 281.
2. Almerico, A. M.; Diana, P.; Barraja, P.; Dattolo, G.; Mingoia, F.; Putzolu, M.; Perra, G.; Milia, C.; Musiu, C.; Marongiu, M. E. *Il Farmaco* **1997**, *52*, 667.
3. Almerico, A. M.; Diana, P.; Barraja, P.; Dattolo, G.; Mingoia, F.; Loi, A. G.; Scintu, F.; Milia, C.; Puddu, I.; La Colla, P. *Il Farmaco* **1998**, *53*, 33.
4. Almerico, A. M.; Barraja, P.; Diana, P.; Cirrincione, G.; Mingoia, F.; Musiu, C.; Perra, G.; Putzolu, M.; Pinna, N.; La Colla, P. *Il Farmaco* **1998**, *53*, 409.
5. All the calculations were run on a Silicon Graphics Indigo II workstation by using the softwares TSAR 3.2 (Tools for Structure Activity Relationships), VAMP 6.0, ASP 3.2 (Automated Similarity Packages) by Oxford Molecular – Accelrys.
6. Zupan, J.; Gasteiger, J. *Neural Networks for Chemists: An Introduction*; VCH: Weinheim, 1993.
7. Molecular descriptors were derived according to the method and assumptions reported in TSAR 3.2 Reference Guide, 1998, Oxford Molecular Limited.
8. More detailed procedure can be found in the TSAR Reference and User guides, 1998, Oxford Molecular Limited.