Pergamon

Bioorganic & Medicinal Chemistry Letters 10 (2000) 2075–2078

BIOORGANIC & MEDICINAL CHEMISTRY LETTERS

A QSAR Study Investigating the Effect of L-Alanine Ester Variation on the Anti-HIV Activity of Some Phosphoramidate Derivatives of d4T

Michael H. Knaggs,[a] Christopher McGuigan,[a] Sarah A. Harris,[a] Parissa Heshmati,[a] Dominique Cahard,[a] Ian H. Gilbert[a,*] and Jan Balzarini[b]

[a]*Welsh School of Pharmacy, Cardiff University, King Edward VII Avenue, Cardiff CF10 3XF, UK*
[b]*Rega Institute for Medical Research, Katholieke Universiteit, Leuven, B-3000 Leuven, Belgium*

Received 14 February 2000; revised 26 June 2000; accepted 6 July 2000

Abstract—A QSAR study, involving the use of calculated physical properties (TSAR™), and the use of a neural network approach (TSAR™), has been performed concerning the anti-HIV activity and cytotoxic effects of a series of d4T phosphoramidate derivatives with varying L-alanine esters. Models were obtained which allow reliable predictions for the anti-HIV activity, and cytotoxicity, of these derivatives.

Introduction

Nucleoside analogues such as d4T and AZT are important drugs for the treatment of HIV infection. These compounds are actually prodrugs, being activated to the nucleoside triphosphates intracellularly by host kinases.[1] Unfortunately this requirement for activation limits the use of some nucleosides. Bypass of any of the steps of activation should increase the therapeutic utility of nucleoside analogues and widen the available choice of potential nucleoside structures. We have reported the synthesis and anti-retroviral activity of the masked alaninyl phosphoramidate of 2′,3′-didehydro-2′,3′-dideoxythymidine (d4T) as a prodrug of the 5′-monophosphate (d4T-MP) (Figure 1 where R = methyl), and showed that some compounds retain full anti-HIV activity in thymidine-kinase deficient CEM/TK$^-$ cells.[2,3,4] These findings indicate successful intracellular nucleotide delivery, and the bypass of the first activation step of d4T by thymidine kinase.

These prodrugs are believed to undergo a partial hydrolysis by carboxyesterase enzymes to novel intermediates, such as alaninyl d4T monophosphophosphate, which may act as depot forms of d4T monophosphate.[3,4] A variety of analogues have been synthesised,[5] and evaluated for anti-HIV activity, with modifications at the L-alaninyl carboxy terminus: these probe the steric and electronic factors at this region, and include primary, secondary, tertiary alkyl, and aryl esters.

Modelling

In an attempt to optimise the conduct of future synthetic work in this series, these phosphoramidate derivatives, for which anti-HIV activity data are available, were subjected to a QSAR (quantitative structure–activity relationship) analysis, in order to determine the physiochemical properties which appear to influence biological activity.[6,7] The activity, log ($1/EC_{50}$) (where EC_{50} is the effective compound concentration (in μM) required to inhibit HIV-induced giant cell formation in CEM cell cultures by 50%) and toxicity, CC_{50} (where CC_{50} is the cytotoxic compound concentration (in μM) required to inhibit CEM cell proliferation by 50%) values were used as the dependent values in the QSAR study (Table 1). QSAR equations were derived for both HIV-1 and HIV-2.

QSAR model for anti-HIV-1 activity (Fig. 2(a))

$$\mathrm{Log}(1/EC_{50}) = -0.0565(X_1)^3 - 0.1745(X_2)^3 - 0.0118(X_3)^3 + 2.849$$

*Corresponding author. Fax: +44-29-2087-4537; e-mail: gilbertih@cf.ac.uk

0960-894X/00/$ - see front matter
PII: S0960-894X(00)00397-8

Table 1. Anti-HIV-1, HIV-2 and cytotoxicity values of compounds in CEM cells[a]

R	log (1/EC_{50}) HIV-1	log (1/EC_{50}) HIV-2	CC_{50}
Methyl	1.13	1.13	100
Ethyl	1.00	1.16	55
n-Propyl	1.33	1.10	40
n-Butyl	NI[b]	NI	60
n-Pentyl	1.35	1.10	40.5
n-Hexyl	1.22	1.26	48
tert-Butyl	0.071	−0.041	250
(*tert*-Butyl)methyl	1.30	1.35	43.5
Cyclohexyl	1.19	1.13	51.4
Cyclohexylmethyl	1.40	NI	12.1
Phenyl	1.60	1.10	NI
Benzyl	1.80	1.80	NI
Phenylethyl	0.071	0.155	250
Phenylpropyl	0.00	−0.217	250
1-Napthylmethyl	1.52	1.22	21.6

[a]EC_{50} and CC_{50} values are in μM.
[b]NI = not included in the final model; data point is an out-lier.

Where:

X_1 = {Verloop B1 (R)}. Covariance standard error value of coefficient = 0.0156
X_2 = {Bond dipole moment (R)}. Covariance standard error value of coefficient = 0.0421
X_3 = {Lipole X component (R)}. Covariance standard error value of coefficient = 0.00161

The statistics[8] for this equation are: $n = 14$, $r = 0.942$, $r^2 = 0.887$, $r^2(\text{cv}) = 0.538$, $s = 0.225$, F probability = 8.64×10^{-6}.

QSAR model for anti-HIV-2 activity (Fig. 2(b))

$$\text{Log}(1/EC_{50}) = -1.1799X_1 - 0.0648(X_2)^3 - 0.0113(X_3)^3 + 3.799$$

Where:

X_1 = Bond dipole moment (R). Covariance standard error value of coefficient = 0.5051
X_2 = {Verloop B1 (R)}. Covariance standard error value of coefficient = 0.0187
X_3 = {Lipole X component (R)}. Covariance standard error value of coefficient = 0.0019

The statistics for this equation are: $n = 13$, $r = 0.918$, $r^2 = 0.843$, $r^2(\text{cv}) = 0.363$, $s = 0.270$, F probability = 1.68×10^{-4}.

The Verloop B1 parameter and lipole X component of the ester substituent are highly correlated with log (1/EC_{50}) for both HIV-1 and HIV-2. Verloop B1 correlates with $r = 0.47$ versus both HIV-1 and HIV-2, and the lipole X component correlates with $r = 0.60$ versus HIV-1 and $r = 0.62$ versus HIV-2.

The Verloop B1 describes the smallest width of the substituent perpendicular to the maxium length of the substituent along the axis of the bond between the first atom of the substituent and the parent molecule. The Lipole X is a measure of the lipophilic distribution along the axis defined by the bond to the point of attachment. Large lipole X values indicate a large distribution of lipophilic groups distant from the point of attachment.

QSAR model for cytotoxicity (Fig. 3)

A QSTR (quantitative structure–toxicity relationship) model was obtained for describing CC_{50}:

$$CC_{50} = -36.296X_1 + 1.3039X_2 + 33.7105X_3 + 113.50726X_4 - 768.3638X_5 + 425.8620X_6 - 6.1983X_7 + 1382.4998$$

Where:

X_1 = Verloop B2 (R). Covariance standard error value of coefficient = 5.5687
X_2 = Ellipsoidal volume (R). Covariance standard error value of coefficient = 0.05237
X_3 = Dipole moment X component (R). Covariance standard error value of coefficient = 9.0475
X_4 = Kappa 2 shape index (R). Covariance standard error value of coefficient = 6.7824
X_5 = Verloop B1 (R). Covariance standard error value of coefficient = 82.753
X_6 = Lipole Z component (R). Covariance standard error value of coefficient = 34.967
X_7 = Bond lipole (R). Covariance standard error value of coefficient = 0.9261

The statistics for this equation are: $n = 13$, $r = 0.999$, $r^2 = 0.998$, $r^2(\text{cv}) = 0.972$, $s = 5.736$, F probability = 1.86×10^{-13}.

The dipole moment X component ($r = 0.72$), the ellipsoid volume ($r = 0.37$), and the Verloop B1 parameter ($r = 0.278$) are the three highest correlating variables with the CC_{50}. The dipole moment X component is analogous to the lipole moment X value except that charges are used instead of lipophilicities. The ellipsoidal volume dependence and Verloop B1 dependence suggest that there is an optimal shape for the substituent, whilst there appears to be an inverse relationship for dipole moments between the HIV-1 and HIV-2 models and the CC_{50} models.

Neural net analysis

In an attempt to improve the predictions obtained for the HIV-1 and HIV-2 data in the above QSAR models, a neural net analysis has been performed using the functionality offered by the TSAR software. The multiple-layer forward feed neural network functionality has been used, which undergoes a supervised training by the back propagation of errors.[9] The inputs for the neural network in this case were the descriptors obtained in the QSAR models described above, while the outputs were the log(1/EC_{50}) values. The neural network functionality within the TSAR software automatically computes

the number of hidden neurons, as well as the number of training and test patterns (in this case 90% used for training and 10% used for test purposes), in order to achieve the best ρ factor,[10] which is the ratio of the number of data points in the training set to the number of variables controlled by the neural network, and in this case ρ had a value of 3.7.

The statistics obtained for the neural net treatment of the HIV-1 data were $n=14$, input columns = 3, net configuration = 3-2-1 (3 input modes, 2 processing nodes, 1 output node), with Test root mean square fit: 0.11 (Fig. 4(a)). The statistics obtained for the neural net treatment of the HIV-2 data were $n=13$, input columns = 3, net configuration = 3-2-1, with Test root mean square fit: 0.03 (Fig. 4(b)).

The QSAR/QSTR models obtained exhibit strong dependencies on the directional component to lipophilicity (lipole moments) and dipole moment effects of the ester substituent. The neural network treatment appears to improve the predictions obtained using the same descriptors used for the HIV-1 and HIV-2 QSAR models. The toxicity model shows that the single most correlated component to CC_{50} is the dipole moment X-component of the ester substituent ($r=0.72$).

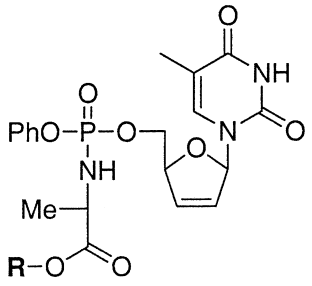

Figure 1.

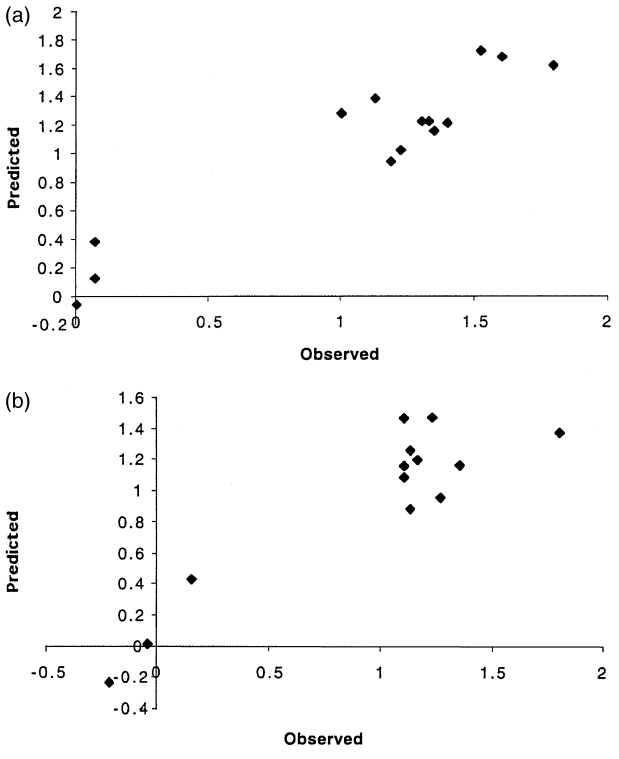

Figure 2. Observed anti-HIV activity plotted against the predicted anti-HIV activity from TSAR. Anti-HIV activity is defined as log ($1/EC_{50}$) where the EC_{50} value is in μM. (a) for HIV-1; (b) for HIV-2.

Discussion

The QSAR/QSTR models obtained for the L-alanine ester derivatives both show a dependence on the molecular shape of the ester substituent. Since the first activation

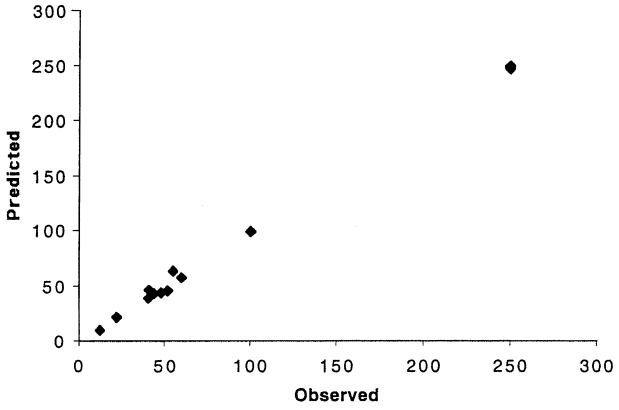

Figure 3. Observed toxicity plotted against predicted toxicity. Toxicity is defined as CC_{50} and is in μM.

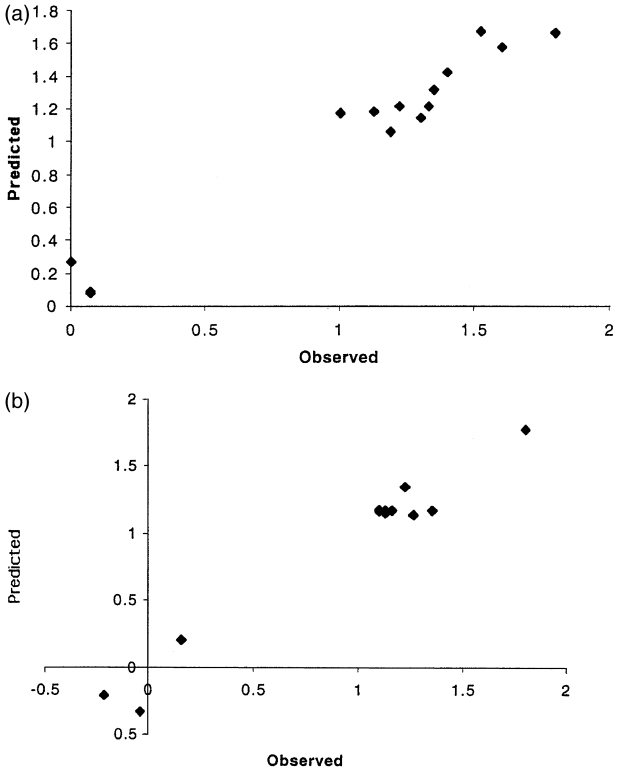

Figure 4. Observed anti-HIV activity plotted against the predicted anti-HIV activity from neural network analysis. Anti-HIV activity is defined as $\log(1/EC_{50})$ where the EC_{50} value is in μM. (a) for HIV-1; (b) for HIV-2.

step for these pro-nucleotides is believed to be carboxyesterase mediated cleavage of the ester substituent, then the correlation of activity and toxicity to this parameter is not unexpected, and may further support the proposed activation of these derivatives.

The QSAR study indicates that for improved anti-HIV activity the Verloop B1 parameter should be minimised, as should the lipole X component and the bond dipole moment for the substituent. This suggests that the incorporation of electronegative groups within the substituent, but distant from the attachment point should increase the activity of these prodrugs. Unfortunately, these same changes are likely to result in compounds which the QSTR study suggests would increase cytotoxicity. However it is likely that optimal values of these common descriptors in both the QSAR and QSTR models exist. It is envisaged that the models obtained will provide a valuable predictive tool for further investigation into the effects of ester variation on the activity and toxicity of d4T phosphoramidate derivatives.

Acknowledgements

We would like to thank Cardiff University and the Cardiff Centre for Molecular Modelling (MHK), and the BBSRC (SAH) for support.

References and Notes

1. Balzarini, J.; Herdewijn, P.; De Clercq, E. *J. Biol. Chem.* **1989**, *264*, 6127. Gao, W.-Y.; Agbaria, R.; Driscoll, J. S.; Mitsuya, H. *J. Biol. Chem.* **1994**, *269*, 12633.
2. McGuigan, C.; Cahard, D.; Sheeka, H. M.; De Clercq, E.; Balzarini, J. *Bioorg. Med. Chem. Lett.* **1996**, *6*, 1183.
3. Balzarini, J.; Egberinck, H.; Hartmann, K.; Cahard, D.; Vahlenkamp, T.; Thormar, H.; De Clercq, E.; McGuigan, C. *Mol. Pharmacol.* **1996**, *50*, 1207.
4. Balzarini, J.; Karlsson, A.; Aquaro, S.; Perno, C.-F.; Cahard, D.; Naesens, L.; De Clercq, E.; McGuigan, C. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 7295.
5. The synthesis and anti-HIV data of compounds is described in the following reference: McGuigan, C.; Tsang, H.-W.; Sutton, P. W.; De Clercq, E.; Balzarini, J. *Antiviral Chem. Chemother.* **1998**, *9*, 109. Compounds where R = ethyl, R = propyl and R = butyl (which are not reported here) were prepared and assayed as described in this reference (R = ethyl: 89% yield, IC_{50} (HIV-1) 0.1 μM; IC_{50} (HIV-2) 0.07 μM; CC_{50} 55 μM. R = propyl: 23% yield, IC_{50} (HIV-1) 0.047 μM; IC_{50} (HIV-2) 0.08 μM; CC_{50} 39.8 μM. R = *n*-butyl: 9% yield; CC_{50} 67.5 μM.).
6. Kubinyi, H. *Drug Discovery Today* **1997**, *2*, 457. Kubinyi, H. *Drug Discovery Today* **1997**, *2*, 538.
7. All structures were generated and minimised, within Sybyl 6.5, using the Tripos force field, with Gasteiger–Huckel charges, and the Powell gradient minimisation technique until an energy gradient of 0.001kcal/mol was obtained. The structures were further optimised, within Sybyl 6.5, using semi-empirical quantum mechanics, utilising the PM3 Hamiltonian with numerical precision. The resulting structures were imported into the TSAR™ package, and for each whole molecule and substituent (R), separately, 60+ electronic, shape and topological descriptors were generated. The least significant substituent and whole molecule descriptors, for both the activity and toxicity models, were determined and rejected using the partial least squares method (PLS) within TSAR. The models were further optimised using multiple regression analysis, within TSAR, in an attempt to reduce the co-linearity of the descriptors, and further rejection of the least significant descriptors in the model based on their *t*-probability values. Due to the small data set, the cross validation technique employed throughout this work was 'leave one out' validation.
8. *n* = number of data points, *r* = correlation coefficient, r^2(cv) = cross-correlated correlation coefficient squared, *s* = standard error.
9. Zupan, J.; Gasteiger, J. *Neural Networks for Chemists: An Introduction*; VCH: Weinheim, 1993.
10. So, S. S.; Richards, W. G. *J. Med. Chem.* **1991**, *35*, 3201.