J Mol Model (2006) 12: 398–405
DOI 10.1007/s00894-005-0059-x

ORIGINAL PAPER

Mohamed Zahouily · Mohamed Lazar · Abdelhakim Elmakssoudi Jamila Rakik · Sanaa Elaychi · A. Rayadh

QSAR for *anti*-malarial activity of 2-aziridinyl and 2,3-bis (aziridinyl)-1,4-naphthoquinonyl sulfonate and acylate derivatives

Received: 25 January 2005 / Accepted: 22 July 2005 / Published online: 9 December 2005

Abstract Quantitative structure-activity antimalarial relationships have been studied for 63 analogues of 2-aziridinyl and 2,3-bis(aziridinyl)-1,4-naphthoquinonyl sulfonate and acylate derivatives by means of multiple linear regression (MLR) and artificial neural networks (ANN). The antimalarial activity [$-\log(IC_{50}\times10^6)$] of the compounds studied were well correlated with descriptors encoding the chemical structure. Using the pertinent descriptors revealed by a stepwise procedure in the multiple linear regression technique, a correlation coefficient of 0.9394 (s=0.2121) for the training set was obtained for the ANN model in a [3-5-1] configuration. The results show that the antimalarial activity of 2-aziridinyl and 2,3-bis(aziridinyl)-1,4-naphthoquinonyl sulfonate and acylate derivatives is strongly dependent on hydrophobic character, hydrogen-bond acceptors and also steric factors of the substituents.

Keywords 2D-QSAR · *Anti*-malarial activity · Pertinent descriptors · MLR · ANN

Introduction

Malaria is serious a sometimes fatal disease caused by a parasite. There are an estimated 300–500 million cases of malaria each year resulting in over 1–2 million deaths [1]. The spread of drug-resistant parasites and the problems associated in controlling vectors are responsible for the continuous rise in the incidence of malaria infection [2]. Although there has been an emphasis on vaccine development [1–8], this has not yet made a significant contribution to controlling the disease. New drugs [9], which are effective against the resistant *Plasmodium falciparum* are sought. Although the discovery of artemisinin, an endoperoxide sesquiterpene lactone and a number of its analogues including trioxane dimer 1,2,4-trioxane have shown high antimalarial activity [10–15] against *P. falciparum*, still the search for antimalarial drugs with increased half-lives and minimum side effects is of current interest [16–23].

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 [24] are located in industrial research units.

The overall picture that emerges from previous QSAR and 3D-QSAR studies shows that the hydrophobic and principally the steric characteristics of substituents have a predominant role in the anti-malarial activity of the set of 2-aziridinyl and 2,3-bis(aziridinyl)-1,4-naphthoquinonyl sulfonate and acylate derivatives [25].

In the present work, we use a combination of multiple linear regression (MLR) and artificial neural network (ANN) techniques for modeling the observed anti-malarial activity of 63 2-aziridinyl and 2,3-bis(aziridinyl)-1,4-naphthoquinonyl sulfonate and acylate derivatives (Fig. 1). The pertinent variable descriptors selected by the first method are introduced as input neurons in the ANN architecture to optimize their non-linear combinations.

Materials and methods

Biological data

The chemical structures along with observed activity data of the compounds used in this study are shown in Table 1. The activity data were taken from various studies [25].

Descriptors

The main step in SAR and QSAR consists in parameterizing the variation in chemical structure. It is obvious that the

M. Zahouily (✉) · M. Lazar · A. Elmakssoudi · J. Rakik · S. Elaychi · A. Rayadh
UFR Chimie Appliquée, Laboratoire de Catalyse, Chimiométrie et Environnement,
Département de Chimie, B.P. 146,
20650 Mohammadia, Maroc
e-mail: mzahouily@yahoo.fr
Fax: +212-23315353

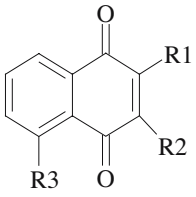

Fig. 1 General structure of 2-Aziridinyl and 2,3-bis (aziridinyl)-1,4-naphthoquinonyl sulfonate and acylate derivatives

performance of QSAR models depends mostly on the parameters used to describe the molecular structures.

In this study, a set of descriptors related to physico-chemical and geometric properties of the molecules was used. In order to study their influence on the in vitro activity of these compounds. All these descriptors were calculated for the separate substituents R_1, R_2 and R_3 (Fig. 1).

Molecular properties used for each substituent, were:

- Size and shape described by means of van der Waals volume (V) and surface (S).
- Molecular dimensions (length, width and height). Length (L): is the distance along the screen x-axis between the left and rightmost atoms plus their van der Waals radii. Width (W): is the distance along the screen y-axis between the top and bottommost atoms plus their van der Waals radii. Height (H): is the distance along the screen z-axis between the nearest and farthest atoms plus their van der Waals radii.
- Ratios V/L, V/W, W/H were also calculated.

Table 1 Chemical structure of *anti*-malarial activity of the set of 2-Aziridinyl and 2,3-bis(aziridinyl)-1,4-naphthoquinonyl sulfonate and acylate derivatives and observed *anti*-malarial activities

No.	R_1	R_2	R_3	$-\log (IC_{50}.10^6)$
1	H	H	OH	−0.5700
2	CH_3	H	OH	−0.3800
3	CH_3	H	H	−0.5300
4	OH	H	H	−2.0000
5	Aziridin−1-yl	H	OH	1.0500
6	Aziridin-1-yl	Aziridin-1-yl	OH	−0.2000
7	CH_3	Aziridin-1-yl	OH	−0.1800
8	$NHCH_3$	H	OH	−0.5700
9	N $(CH_3)_2$	H	OH	−0.2000
10	NH $(CH_2)_2Cl$	H	OH	−0.6300
11	Aziridin-1-yl	H	$C_6H_5SO_3$	0.6400
12	Aziridin-1-yl	H	CH_3-4-$C_6H_4SO_3$	0.7000
13	Aziridin-1-yl	H	C_2H_5-4-$C_6H_4SO_3$	1.0200
14	Aziridin-1-yl	H	$(CH_3)_3C$-4-$C_6H_4SO_3$	1.6200
15	Aziridin-1-yl	H	F-4-$C_6H_4SO_3$	0.8000
16	Aziridin-1-yl	H	CH_3O-4-$C_6H_4SO_3$	0.8500
17	Aziridin-1-yl	H	$C_6H_5CH{=}HSO_3$	0.7700
18	Aziridin-1-yl	H	$(CH_3)_2CH$-2, 4,6-$C_6H_2SO_3$	0.5500
19	Aziridin-1-yl	H	Naphtaléne-1-sulfonyloxy	0.8900
20	Aziridin 1 yl	H	$CH_3)_2N$ 5 naphtaléne1 sulfonyloxy	0.5400
21	Aziridin 1 yl	H	Quinolin 8 sulfonyloxy	0.3700
22	Aziridin 1 yl	H	Thiophene 2 sulfonyloxy	0.3300
23	Aziridin 1 yl	Aziridin 1 yl	$C_6H_5SO_3$	−0.7600
24	Aziridin-1-yl	Aziridin-1-yl	CH_3-4-$C_6H_4SO_3$	−0.6600
25	Aziridin-1-yl	Aziridin-1-yl	C_2H_5-4-$C_6H_4SO_3$	−0.6000
26	Aziridin-1-yl	Aziridin-1-yl	$(CH_3)_3C$-4-$C_6H_4SO_3$	−0.6600
27	Aziridin-1-yl	Aziridin-1-yl	Cl-4-$C_6H_4SO_3$	−0.5000
28	Aziridin-1-yl	Aziridin-1-yl	Br-4-$C_6H_4SO_3$	−0.5700
29	Aziridin-1-yl	Aziridin-1-yl	NO_2-4-$C_6H_4SO_3$	−0.8800
30	Aziridin-1-yl	Aziridin-1-yl	NO_2-3,4-$C_6H_4SO_3$	−0.3800
31	Aziridin-1-yl	Aziridin-1-yl	CH_3-2, 4,6-$C_6H_2SO_3$	−0.5900
32	Aziridin-1-yl	Aziridin-1-yl	$(CH_3)_2CH$-2, 4,6-$C_6H_2SO_3$	−0.4800
33	Aziridin-1-yl	Aziridin-1-yl	$C_6H_5CH_2SO_3$	−0.3800
34	Aziridin-1-yl	Aziridin-1-yl	$C_6H_5CH{=}CHSO_3$	−0.7600
35	Aziridin-1-yl	Aziridin-1-yl	Naphtaléne-1-sulfonyloxy	−0.5300
36	Aziridin-1-yl	Aziridin-1-yl	$CH_3)_2N$-5-naphtaléne1-sulfonyloxy	−0.4800
37	Aziridin-1-yl	Aziridin-1-yl	CH_3	−0.5000

Table 1 (continued)

No.	R_1	R_2	R_3	$-\log (IC_{50}.10^6)$
38	Aziridin-1-yl	Aziridin-1-yl	CH_3 $(CH_2)_{14}CH_2SO_3$	−1.3000
39	Aziridin-1-yl	Aziridin-1-yl	Cl (CH_2) CH_2SO_3	−0.4800
40	Aziridin-1-yl	Aziridin-1-yl	{Bicyclo[2,2,1]hepta-7-dimetyl-2-one}methyl-sulfonyloxy	−0.7800
41	Aziridin-1-yl	H	$C_6H_5CO_2$	0.3500
42	Aziridin-1-yl	H	CH_3-4-$C_6H_4CO_2$	0.2100
43	Aziridin-1-yl	H	F-4-$C_6H_4CO_2$	0.3100
44	Aziridin-1-yl	H	Cl-4-$C_6H_4CO_2$	0.2800
45	Aziridin-1-yl	H	CH_3O-4-$C_6H_4CO_2$	0.2700
46	Aziridin-1-yl	H	CH_3O-3, 4,6-$C_6H_2CO_2$	0.8000
47	Aziridin-1-yl	H	Furan-2-carbonyloxy	0.3100
48	Aziridin-1-yl	H	Thiophene-2-carbonyloxy	0.3300
49	Aziridin-1-yl	Aziridin-1-yl	$C_6H_5CO_2$	−0.4100
50	Aziridin-1-yl	Aziridin-1-yl	CH_3-4-$C_6H_4CO_2$	−0.6400
51	Aziridin-1-yl	Aziridin-1-yl	F-4-$C_6H_4CO_2$	−0.6400
52	Aziridin-1-yl	Aziridin-1-yl	Cl-4-$C_6H_4CO_2$	−0.2300
53	Aziridin-1-yl	Aziridin-1-yl	CH_3O-4-$C_6H_4CO_2$	−0.5700
54	Aziridin-1-yl	Aziridin-1-yl	CH_3O-3, 4,6-$C_6H_2CO_2$	−0.5300
55	Aziridin-1-yl	Aziridin-1-yl	Furan-2-carbonyloxy	−0.6600
56	Aziridin-1-yl	Aziridin-1-yl	Thiophene-2-carbonyloxy	−0.4100
57	Aziridin-1-yl	Aziridin-1-yl	2,5-dichloro-thiophene3-carbonyloxy	−0.5900
58	Aziridin-1-yl	Aziridin-1-yl	Adamantan-1-carbonyloxy	−0.7300
59	Cl $(CH_2)_2NH_2$	H	CH_3SO_3	−0.9300
60	Cl $(CH_2)_2NH_2$	H	$(CH_3)_2NSO_3$	−1.3400
61	Cl $(CH_2)_2NH_2$	H	$C_6H_5CH_2SO_3$	−0.8100
62	Cl $(CH_2)_2NH_2$	H	$(CH_3)_2$N-5-Naphtaléne1-sulfonyloxy	−0.4600
63	CH_3	Aziridin-1-yl	$(CH_3)_3$C-4-$C_6H_4SO_3$	−0.5600

– log *P*, the partition coefficient between *n*-octanol and water.
– Molar refractivity (MR).
– Molecular weight (MW).
– Hydrogen-bonding donors (HBD) and hydrogen-bonding acceptors (HBA).
– Electronegativity.

All these descriptors were calculated with the demo version of the molecular modeling program (MMP) [26a].

The topological descriptors for each substituent are the connectivity indices from Kier and Hall [27] "up to fourth order", and the electrotopological indices [28].

Statistical analysis

2D-QSAR models were derived using multiple linear regression [26b] and artificial neural networks [26c]. The predictivity potential of the models was determined by cross-validation methods [29].

Table 2 Descriptive statistics of 63 compounds

	Avg*[a]	STD[b]	Var[c]	Sum[d]	SSQ[e]	Min[f]	Max[g]
Log $(IC_{50}.10^6)_{Obs}$	−0.2075	0.552	0.456	−13.07	28.252	−2.00	1.62
Log $(IC_{50}.10^6)_{Cal}$	−0.2223	0.475	0.2906	−14.0039	18.0169	−1.2786	0.7674
Log $P(R_2)$	−0.167	0.173	0.030	−10.705	1.878	−0.345	0.000
MW (R_3)	153.253	59.985	6045.904	9654.942	374846.1	1.008	305.502
HBA (R_1)	0.0420	0.074	0.017	2.647	1.086	0.000	0.524

[a]Avg*: mean value;
[b]STD: standard deviation;
[c]Var: variance;
[d]Sum: sum of values;
[e]SSQ: sum of Square,
[f]Min: minimum,
[g]Max: maximum

Table 3 Correlation matrix of data

	$-\log (1/IC_{50} \cdot 10^6)$	HBA (R_1)	Log P (R_2)	Mw (R_3)
$-\log (1/IC_{50} \cdot 10^6)$	1.0000			
HBA(R_1)	0.6642	1.0000		
LogP(R_2)	−0.7720	−0.2374	1.0000	
Mw(R_3)	−0.5961	−0.3261	0.7288	1.0000

The multiple linear regression method was used to generate linear models between the antimalarial activity and the molecular descriptors.

Because of the large number of descriptors considered, a stepwise procedure combining the forward and backward algorithms was used to select the pertinent descriptors.

In order to avoid all difficulties in the interpretation of the resulting models, pairs of variables with a correlation coefficient larger than 0.80 were classified as inter-correlated, and only one of these was included in the screened model. The quality of the model was considered as statistically satisfactory on the basis of squared correlation coefficient (r^2), standard deviation (s), and F-statistics (F) when all parameters in the model were significant at the 95% confidence level ($p<0.05$).

The application of artificial neural networks (ANNs) to solving problems in chemistry is a recent research field. ANN have been used to the investigate quantitative structure activity relationships (QSAR) [30, 31].

Neural network models are known to be very effective in representing the non-linear relationships which could exist between variables in complex systems. For most applications of ANNs to chemistry, ANNs using the back propagation algorithm (BPA), which is used in this study, seems to be a good choice [32].

Results and discussion

Multiple linear regression analysis

Regression equations

All the compounds were considered for the development of 2D-QSAR models between the physicochemical and electrotopology parameters as independent and $-\log(IC_{50}\times10^6)$ value against p. falciparum as the independent variable.

Table 5 Descriptors contribution in Eq. 3

Descriptors	Contribution (%)
Log $P(R_2)$	47.64
MW (R_3)	18.63
HBA (R_1)	33.73

In this work we used centered and reduced values [33], calculated according to Eq. 1, of relevant descriptors selected, in order to have homogeneity in our data.

$$d_{ij} = \frac{\left(Y_{ij} - \overline{Y_j}\right)}{\sqrt{Var_j}} \tag{1}$$

d_{ij}: Centered and reduced value of descriptor j.

Y_{ij}: Values of descriptor j for each molecule i.

*$\overline{Y_j}$: Average of values of descriptor j.

*$\sqrt{Var_j}$: Root Square of variance for descriptor j.

After collecting the data, we submitted all parameters to the regression; many models were generated using this method. We obtained the best models without constant terms (Eq. 3) because the constant term it not statistically significant. However, an ideal model (Eq. 3) is one that has high r^2 and F values, low standard deviation, least numbers of independent variables, and high ability for prediction.

$$\begin{aligned} -\log\left(1/IC_{50}\times10^6\right) &= 0.0554(\pm0.4019) - 2.8206(\pm0.4231)HBA(R_1) \\ &\quad + 3.0120(\pm0.3314)\log P(R_2) \\ &\quad + 0.0024(\pm0.0007)\,MW(R_3) \end{aligned}$$

$$n = 63 \quad r = 0.79\left(r^2 = 0.63\right) \quad s = 0.2772 \quad F-\text{ratio} = 34.11 \tag{2}$$

$$\begin{aligned} -\log\left(1/IC_{50}\times10^6\right) &= -2.7660(\pm0.4019)\,HBA(R_1) \\ &\quad + 2.9705(\pm0.3157)\log P(R_2) \\ &\quad + 0.0026(\pm0.0004)MW(R_3) \end{aligned}$$

Table 4 Descriptive statistics of 58 compounds

	Avg[a]	STD	Var	Sum	SSQ	Min	Max
Log $(IC_{50}\cdot10^6)_{Obs}$	−0.211	0.495	0.332	−12.260	18.923	−1.340	1.02
Log $(IC_{50}\cdot10^6)_{Cal}$	−0.223	0.481	0.294	−12.945	16.728	−1.279	0.767
Log $P(R_2)$	−0.173	0.173	0.030	−10.015	1.729	−0.3453	0.000
MW (R_3)	156.916	53.382	5034.749	9101.141	286980.689	1.008	283.412
HBA (R_1)	0.043	0.077	0.0187	2.485	1.064	0.000	0.524

[a]For the significance, see Table 2

$$n = 63 \quad r = 0.81\left(r^2 = 0.66\right) \quad s = 0.2775 \quad F\text{-ratio} = 39.74 \tag{3}$$

The statistical quality of Eq. 3 is not satisfactory and accounts for 66% of the variance in $-\log(1/IC_{50}\times10^6)$ but a standard deviation inferior to that associated with the mean value of $-\log(1/IC_{50}\times10^6)$ (see Table 2).

To remain close to the experimental error (5%), we take away the five molecules having $di = \frac{|obs_i - cal_i|}{obs_i}$ higher than $2s/(\log 1/EC_{50})$. Where $\overline{\log 1/EC_{50}}$ is the mean of observed activity. Consequently, a new regression model was derived using 58 molecules (Eq. 4).

$$-\log\left(1/IC_{50}\times10^6\right) = -2.5315(\pm0.2643)\mathrm{HBA}(R_1) + 2.9102(\pm0.2140)\log \mathrm{P}(R_2) + 0.0026(\pm0.0003)\mathrm{MW}(R_3)$$

$$n = 58 \quad r = 0.90\left(r^2 = 0.82\right) \quad s = 0.2040 \quad F\text{-ratio} = 81.35 \tag{4}$$

This equation shows that the regression coefficients of hydrogen-bond acceptors of the substituent R_1 followed by the hydrophobic character of substituents expressed by log $P(R_2)$ together have the highest importance.

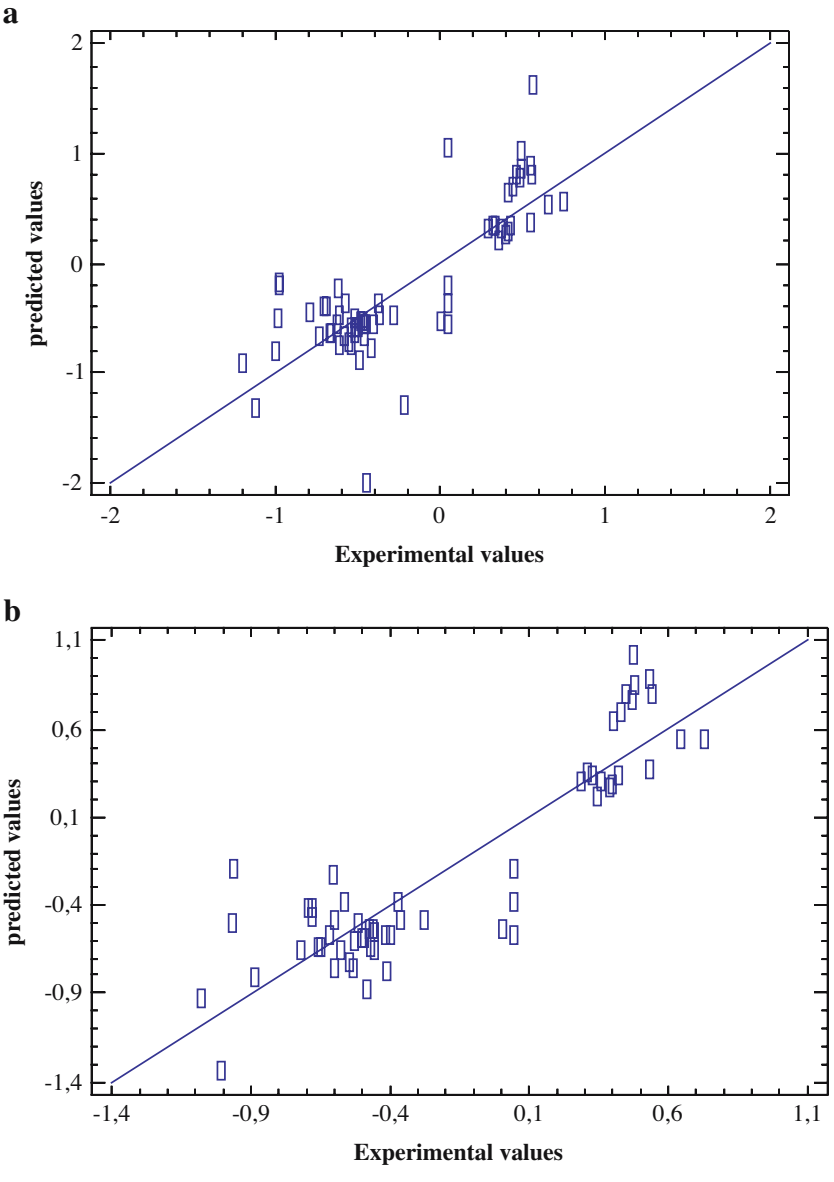

Fig. 2 Observed and predicted values of MLR (**a**) and ANN (**b**) models

Table 6 Variation of r and s with number of hidden neurones

Number of neurones on the hidden layer	s	r
3	0.2387	0.9231
4	0.2132	0.9388
5	0.2121	0.9394
6	0.2133	0.9387
7	0.2151	0.9377
8	0.2147	0.9379

It is noteworthy that there is no significant intercorrelation between the descriptors that appear in the model selected, as seen in Table 3.

Statistical criteria of the model are fairly good. Indeed we have a model with about 82% of the total variance and a standard deviation lower than that associated with the mean value of $-\log (1/IC_{50}\times 10^6)$, as seen in Table 4.

The descriptors' contributions to this equation (see Table 5), calculated according to the Gore method [34], also justified this result.

The large contribution of hydrophobic character of substituents R_2 has a strong effect on the activity. The hydrogen-bond acceptors and also the steric effect of substituents expressed by HBA (R_1) and MW (R_3), respectively, seem to be more significant for the activity. This result is also in agreement with previous ones for this series [35].

The plot in Fig. 2 indicates that there is a significant correlation between actual values and calculated values of $-\log (1/IC_{50} \,.\, 10^6)$ from Eq. 4.

Cross-validation

In the cross-validation phase, 58 subsets were created according to the leave-one-out method and the output of the removed compound was predicted for each subset [36]. A cross-validation coefficient q^2 was calculated according the following equation: [29].

$$q^2 = 1 - (\text{PRESS}/\text{Variance}) \tag{5}$$

Where PRESS is the predictive residual sum of squares. They yielded a q^2=0.77, indicating good predictive quality of the model, according to Wold [37].

Table 7 Evaluating the impact of each descriptor in ANN

Removed descriptor	Ci ANN[a]	R ANN[b]	s ANN[b]
HBA(R_1)	24.77	0.7765	0.3585
logP(R_2)	51.28	0.3996	0.5216
MW(R_3)	23.95	0.8300	0.3452

[a]The contribution (C%) of descriptor given by the second method described in the text
[b]Given by the first method described in the text

Table 8 Cross-validation parameters for the MLR and ANN models

Model	q^2	Q	PRESS	SSY	PRESS/SSY
MLR (58)	0.77	4.4118	4.383	18.9225	0.2316
ANN (58)	0.89	4.4290	2.152	18.9225	0.1137

Artificial neural network analysis

In order to test the possibility of non-linear effects on the data and to establish a more accurate model, we used a neural network technique [30, 31].

The ANN was generated by using the pertinent descriptors appearing in the MLR model as input. A 3-5-1 neural network architecture was developed with the optimum momentum and learning rate of 0.9 and 0.02, respectively and with 10,000 iterations. The five hidden neurons were chosen to maintain ρ [38] between 1.8 and 2.2. To verify this condition, we also tried three to eight neurons in the hidden layer and it was found that five hidden neurons gives the best result for the training and test sets, as shown in Table 6.

To evaluate the neural network, the correlation coefficient r of its results is compared with the r or the regression model developed in this work. The r values were 0.90 and 0.9394 for the training set in the present MLR and the ANN, respectively. The corresponding standard error s for both models was 0.204 and 0.2121, respectively. This preliminary study enables us to conclude that the ANN with the (3-5-1) architecture was able to establish a satisfactory relationship between the pertinent descriptors and activity of 2-aziridinyl and 2,3-bis(aziridinyl)-1,4-naphthoquinonyl sulfonate and acylate derivatives.

Analysis of descriptor's contribution in ANN model

To estimate the relative contribution of descriptors, we have chosen two different approaches:

(i) The contribution of descriptor i (i=1−3) was estimated from the trained 3-5-1 configuration network. The descriptor under study was removed from the 3-5-1 trained ANN together with its corresponding weights. Then the network (3-5-1) calculated the output of each molecule as usual. The mean of the deviations absolute values Δm_i between the observed activity and the estimated activity for

Table 9 Comparison of our models with other 3D-QSAR models

Models	Number of molecules	r	Number of descriptors
Model 1 (Eq. 2)	63	0.81	3
Model 2 (Eq. 4)	58	0.90	3
Model 3 ANN	58	0.939	3
Model 4 [35]	63	0.842	3
Model 5 [35]	62	0.842	4
Model 6 [35]	62	0.853	5

all compounds was calculated. This process was reiterated for each descriptor. Finally, the contribution C_i [38] of descriptor i is given by:

$$C_i = 100.\Delta m_i \Big/ \sum_{i=1}^{4} \Delta m_i \quad (6)$$

(ii) we analyze deviations when a given descriptor is removed and for the full set of descriptors. This approach is an extension of the previous one proposed by Chastrette et al. [39] In that way we could estimate the contribution of each descriptor removed in the model. Table 7 shows that these two methods give qualitatively similar results.

These results indicate that the relative importance of the descriptors varied in the following order: log $P(R_2)$>HBA (R_1)>MW(R_3). Comparison of this classification with the one obtained in the regression reveals a change of value of the contribution of HBA(R_1), and MW(R_3). This could be explained by the possible existence of a non-linear relationship between the activity and hydrogen-bonding acceptors (HBA), which is not the case for the hydrophobicity.

Cross-validation

We used the same procedure as for the MLR analysis and obtained a coefficient of cross-validation equal to q^2=0.89. The model obtained was considered to be predictive according to Wold [37]. The performance of the ANN is superior to that of MLR and this indicates the presence of nonlinearity in the data since the efficiency of the descriptors was increased. The combination of MLR and ANN for descriptor selection was fruitful.

It is worth mentioning that r alone is not the only parameter for deciding the quality of a model. In addition to r, one has to consider the standard error of estimation. In the literature a quality factor Q was introduced [40, 41]. This quality factor, Q, is defined as the ratio of the correlation coefficient r to the standard error of estimation, Se (Q=r/s). That is, the quality of models is judged considering r and s simultaneously. Such Q values are given in Table 8. The Q values reported in Table 8 indicate that the ANN model is better statistically.

To ensure that the results obtained were not due to chance and lend credence to our results, we have run a scrambling experiment [42] and calculated PRESS (Eq. 5) and SSY parameters (SSY is the variance of the biological activity of the molecules around the mean value).

Firstly, the dependent variable ($-\log (IC_{50} \cdot 10^{-6})$ was randomly scrambled and then the same algorithms used in MLR and ANN run once again. The statistical results as the correlation coefficient r and the standard deviation of its results are compared with the r and s of the MLR and ANN models developed in this work. The r values were 0.190 and 0.370 compared with 0.900 and 0.9392 for the s values we have obtained 0.432 and 0.402 compared with 0.2040 and 0.2121 for the training set in MLR and ANN, respectively. This test confirms and clearly shows that the descriptors selected in this study describe the activity studied very well.

PRESS is a good estimate of the real prediction of error of the model, provided that the observations were independent. If PRESS is smaller than the sum of the squares of the response value (SSY), the model predicts better than chance and can be considered statistically significant. Table 8 shows that in with models PRESS is significantly smaller than SSY, indicating them to be statistically significant.

The results obtained by our models are equivalent or better those from other 3D-QSAR models [35] (Table 9).

Conclusion

Taking into account the complexity of the phenomena modeled, we were able to show with few descriptors (three descriptors) and in a 2D QSAR study, that the activity of the 2-aziridinyl and 2,3-bis(aziridinyl)-1,4-naphthoquinonyl sulfonate and acylate derivatives, depends strongly on the steric factors and the hydrophobicity of the substituents attached to the 2-aziridinyl and 2,3-bis(aziridinyl)-1,4-naphthoquinonyl sulfonate and acylate skeletons.

The pattern obtained with the ANN approach is more efficient than regression analysis, since it reveals the non-linear effects in 2-aziridinyl and 2,3-bis(aziridinyl)-1,4-naphthoquinonyl sulfonate and acylate analogues. In addition, the approach used for the contributions and classification of descriptors in the ANN may be of help in QSAR interpretations. The combination between MLR and ANN is revealed as an interesting approach.

References

1. WHO Report (1998) http://www.who.int/inf-fs/en/fact094.html
2. Ghosh A, Edwards MJ, Jacobs-Lorena M (2000) Parasitol Today 16:196-201
3. Macreadie I, Ginsburg H, Sirawaraporn W, Tillry L (2001) Parasitol Today 16:438–443
4. Geary TG, Edgar SA, Jensen JB (1986) In: Campbell WC, Rew RS (eds) Chemotherapy of parasitic diseases. Plenum, New York, pp 209–236
5. Peters W (1985) Parasitology 90:705–715
6. Moran JS, Bernard KW (1989) J Am Med Assoc 262:245–248
7. Wyler DJ, Engl N (1983) J Med 308:875–878
8. Anders RF, Saul A (2000) Parasitol Today 16:444–450
9. Winstanley PA (2000) Parasitol Today 16:146–153
10. Haji H, Smith T, Charlwood JD, Meuwissen JH (1996) Parasitology 113:425–431
11. Abouabdellah A, Begue JP, Bonnet-Delpon D, Gantier JC, Nga TTT, Thac TD (1996) Bioor Med Chem Lett 6:2717–2720
12. Posner GH, Ploypradith P, Parker MH, O'Dowd H, Woo SH, Northrop J, Krasavin M, Dolan P, Kensler TW, Xie S, Shapiro TA (1999) J Med Chem 42:4275–4280
13. Li Y, Zhu YM, Jiang HJ, Pan JP, Wu GS, Shi YL, Yang JD, Wu BA (2000) J Med Chem 43:1635–1640
14. Mekonnen B, Weiss E, Katz E, Ma J, Ziffer H, Kyle DE (2000) Bioorg Med Chem 8:1111–1116
15. Posner GH, Maxwell JP, O'Dowd H, Krasavin M, Xie S, Shapiro TA (2000) Bioorg Med Chem 8:1361–1370

16. O'Neill PM, Miller A, Bishop LPD, Hindley S, Maggs JL, Ward SA, Roberts SM, Scheinmann F, Stachulski AV, Posner GH, Park BK (2001) J Med Chem 44:58–68
17. Jefferies D (1998) Parasitol Today 14:202–206
18. Shahabuddin M, Cociancich S, Zieler H (1998) Parasitol Today 14:493–497
19. Calas M, Ancelin ML, Cordina G, Portefaix P, Piquet G, Vidal-Sailhan V, Vial H (2000) J Med Chem 43:505–516
20. McCullough KJ, Wood JK, Bhattachajee AK, Dong Y, Kyle DE, Milhous WK, Vennerstrom JL (2000) J Med Chem 43:1246–1249
21. Rastelli G, Sinawaraporn W, Sompornpisut P, Vilaivan T, Kamchonwongpaisan S, Quarrell R, Lowe G, Thebtaranonth Y, Yuthavong Y (2000) Bioorg Med Chem 8:1117–1128
22. Karle JM, Bhattacharjee AK (1999) Bioorg Med Chem 7:1769–1774
23. Biot C, Delhaes L, MN'Diaye C, Maciejewski LA, Camus D, Dive D, Brocard JS (1999) Bioorg Med Chem 7:2843–2847
24. Trinajstic N (1992) Chemical graph theory, 2nd edn. CRC, Boca Raton, Florida, p 20
25. Lin TS, Zhu LY, Xu SP, Divo AA, Sartorelli AC (1991) J Med Chem 34:1634–1639
26. (a) MMP, molecular modeling pro-Demo (TM) Revision 301 demo published by ChemSW Software (TM) (b) Unistat statistical package, version 4.0 for Excel (c) Data pro Qnet 2000 for Windows V2 K build neutral network modeling. Vesta Service, Winnetka, III
27. Kier LB, Hall LH (1990) Pharm Res 7:801–807
28. Kier LB, Hall LH (1997) J Chem Inf Comput Sci 37:548–552
29. Tetko IV, Villa AEP, Livingstone DJ (1996) J Chem Inf Comput Sci 36:794–803
30. Zahouily M, Rhihil A, Bazoui H, Sebti S, Zakarya D (2002) J Mol Model 8:168–172
31. Zahouily M, Rayadh A, Aadil M, Zakarya D (2003) J Mol Model 9:242–247
32. Rumhelart DE, Hinton CE, Williams RJ (1986) Nature 323:533–536
33. Bazoui H, Zahouily M, Boulaajaj S, Sebti S, Zakarya D (2002) SAR and QSAR in Environmental Research 13:567–577
34. Gore WL (1952) Interscience. New York, p 141
35. Pandey SK, Naware NB, Trivedi P, Saxena AK (2001) SAR and QSAR in Environmental Research 12:547–564
36. Rumhelart DE, Hinton CE, Williams RJ (1986) Nature 323:533–536
37. Wold S (1991) Quant Struct-Act Relat 10:191–193
38. Cherqaoui D, Esseffar M, Villemin D, Cence JM, Chastrette M, Zakarya D (1998) New J Chem 839–843
39. Chastrette M, Zakarya D, Peyraud JF (1994) Eur J Med Chem 29:343–348
40. Pogliani L (1994) Amino Acids 6:141–153
41. Pogliani L (1996) J Phys Chem 100:18065–18077
42. Bazoui H, Zahouily M, Sebti S, Boulajaaj S, Zakarya D (2002) J Mol Model 8:1–7