J Mol Model (2002) 8:237–245
DOI 10.1007/s00894-002-0088-7

ORIGINAL PAPER

Humberto González Díaz · Ervelio Olazabal
Nilo Castañedo · Ivan Hernádez Sánchez
Alcidez Morales · Hector S. Serrano
Julio González · Ronal Ramos de Armas

Markovian chemicals "in silico" design (MARCH-INSIDE), a promising approach for computer aided molecular design II: experimental and theoretical assessment of a novel method for virtual screening of fasciolicides

Received: 12 March 2002 / Accepted: 4 May 2002 / Published online: 3 August 2002
© Springer-Verlag 2002

Abstract A novel method for in silico selection of fluckicidal drugs is introduced. Two QSARs that permit us to discriminate between fasciolicide and non-fasciolicide drugs (the first) and to outline some conclusions about the possible mechanism of action of a chemical (the second) are performed. The first model correctly classified 93.85% of compounds in the training series and 89.5% of the compounds in the predicting one. This model correctly classified 87.7, 93.8, 92.2 and 93.9% of compounds in leave-*n*-out cross validation procedures when *n* takes values from 2 to until 6. The model seems to be stable in around 92% of good classification in leave-*n*-out cross validation analysis when $n>6$. The second model correctly classified 70% of non-fasciolicide compounds, 85.71% of β-tubulin inhibitors and 100% of proton ionophores in the training set. This model recognizes as proton ionophores 100% of any nitrosalicylanilides in the predicting series. Both models have a low *p*-level <0.05. Finally, the experimental assay of six organic chemicals by an in vivo test permit us to carry out an assessment of the model with a fairly good 100% agreement between experiment and theoretical prediction.

H. González Díaz (✉) · E. Olazabal · N. Castañedo
I.H. Sánchez · A. Morales · H.S. Serrano · J. González
Chemical Bio-actives Center, Central University of "Las Villas", Santa Clara 54830, Villa Clara, Cuba
e-mail: humbertogd@navegalia.com
Tel.: +53/42/281473-131, Fax: +53/42/281473-455

H. González Díaz
Department of Organic Chemistry, Faculty of Pharmacy, University of "Santiago de Compostela" 15706, Spain

R.R. de Armas
Department of Chemistry, Chemistry and Pharmacy Faculty, Central University of "Las Villas". Santa Clara 54830, Villa Clara, Cuba

Keywords Drug design · Stochastic matrix · Markov's chains · QSAR · Fluckicidal drugs · Linear discriminant analysis and electronegativity

Introduction

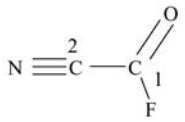
Fasciolosis is a term used to characterize a pathology that some parasite species cause. These parasites are known as Flukes. One of the most common naturally occurring flukes is *Fasciola hepatica*. [1] *Fasciola hepatica* as well as its tropical counterpart *Fasciola gigantic* are widely distributed around the world. Nowadays they are prevalent in South America, the Caribbean region, Europe, and Australia. [2] *Fasciola hepatica* remains one of the single most important helminthes of livestock in the U.K. [3]

The infestation of ruminants with *F. hepatica* causes a significant economic loss, forecasted to be more than U.S.$2,000 million in the world agricultural sector with approximately 600 million infected animals. [4] The WHO estimates that 2.4 million people are infected with fasciolosis and 180 million are at risk of infestation. [5]

A high prevalence of human fasciolosis has been reported in Bolivia, Cuba and Peru, where this illness is recognized as a serious health problem. [6, 7] There are effective strategies for the control of fasciolosis, based on the use of drugs (fasciolicides) together with epidemiological data. [8] Nevertheless, no fasciolicides have been marketed since the 1980s. Consequently, we will have to rely on existing drugs for some time. Thus, we can expect that there will be a great necessity for safer, cheaper, and more active fasciolicide compounds in the near future. [8, 3] On the other hand, computer aided drug design has emerged as a rational alternative in the search for novel drugs. [9, 10] Lajiness [11] and Estrada et al. [12, 13] report a high incidence of the use of novel

Table 1 Definition and calculation of the $^1\prod$ matrix for nitrilo-acetyl fluoride and its molecular structure

Structure:



Calculation:

$$^1\Pi = \begin{array}{c|ccccc} & O & F & C_1 & C_2 & N \\ O & 0.583 & 0 & 0.417 & 0 & 0 \\ F & 0 & 0.615 & 0.385 & 0 & 0 \\ C_1 & 0.280 & 0.320 & 0.200 & 0.200 & 0 \\ C_2 & 0 & 0 & 0.313 & 0.313 & 0.375 \\ N & 0 & 0 & 0 & 0.455 & 0.545 \end{array}$$

Definition[a]:

$$^1\Pi = \begin{array}{c|ccccc} & O & F & C_1 & C_2 & N \\ O & \frac{O}{O+C1} & 0 & \frac{C1}{O+C1} & 0 & 0 \\ F & 0 & \frac{F}{F+C1} & \frac{C1}{F+C1} & 0 & 0 \\ C_1 & \frac{O}{O+F+C1+C2} & \frac{F}{O+F+C1+C2} & \frac{C1}{O+F+C1+C2} & \frac{C2}{O+F+C1+C2} & 0 \\ C_2 & 0 & 0 & \frac{C1}{C1+C2+F} & \frac{C2}{C1+C2+F} & \frac{F}{C1+C2+F} \\ N & 0 & 0 & 0 & \frac{C2}{C2+F} & \frac{F}{C2+F} \end{array}$$

[a] In the definition of the $^1\prod$ matrix, the chemical symbol of the element is used to indicate the corresponding electronegativity value. That is: if we write O it means χ(O), oxygen electronegativity

molecular indices to develop QSAR for in silico virtual drug screening. In this sense the definition of novel molecular descriptors (see Todeschini and Consonni [14] for an exhaustive compilation) is a promising field in medicinal chemistry and veterinarian sciences. Thus, our aims in the present work are: firstly, to fit a classification function that permits us to discriminate between fasciolicide and non-fasciolicide compounds using the MARCH-INSIDE and LDA (linear discriminant analysis) methodologies. [15] Secondly, we shall perform another LDA in order to classify organic chemicals according to their fasciolicide mechanism of action. This analysis will permit us to obtain some estimation about the possible mechanism of action of fasciolicides. Finally, we aim to carry out a primary screening of these compounds and other compounds with similar structure predicted as non-active and to carry out an experimental corroboration of the models. We recall here that the experimental section is aimed at testing experimentally the predictions of the classification function and is not intended to test the real overall effectiveness of the screened drugs. These experiments must be considered only as preliminary screening results.

Materials and methods

Markovian chemicals "in silico" design (MARCH-INSIDE)

The MARCH-INSIDE methodology uses Markov's chain (MCH) [16] to codify information about the molecular structure. This procedure considers the external electron layers of any atom core in the molecule (the valence shell) as states of the MCH. [17] The method uses as source of molecular descriptors the matrix $^1\Pi$, which has the elements p_{ij}. This matrix is called the 1-step electron-transition stochastic matrix. $^1\Pi$ is built as a square table of order n, where n represents the number of atoms in the molecule. The elements ($^1p_{ij}$) of the 1-step electron-transition stochastic matrix are the transition probabilities:

$$^1p_{ij} = \frac{\chi_j}{\sum_{k=1}^{\delta+1} \chi_k} \quad (1)$$

where χ_j is the electronegativity of the atom j, which is bonded with the atom i. [17] The elements of $^1\Pi$ ($^1p_{ij}$) are defined to codify information about the electron-withdrawing strength of atoms to withdraw electrons from their neighbors in the molecule. The MARCH-INSIDE molecular descriptors are defined as:

$$^{SR}\pi_k(S_m) = \sum_{i=1}^{g} {}^k p_{ii} = \mathrm{Tr}\left[\left(^1\pi\right)^k\right] \quad (2)$$

These molecular descriptors are the traces of the kth-step-electron-transition stochastic matrices ($^k\Pi$). These matrices are the successive powers of $^1\Pi$. The trace (Tr) is the sum of the main diagonal elements ($^kp_{ii}$) of $^1\Pi$. [12, 13] In Table 1 the construction of the $^1\Pi$ matrix for nitrilo-acetyl fluoride is exemplified. As could be observed, the p_{ij} values are proportional to the electronegativity of the atom a_j (the atom that attracts the electrons of a_i). Conversely, the p_{ij} values are in inverse relation to

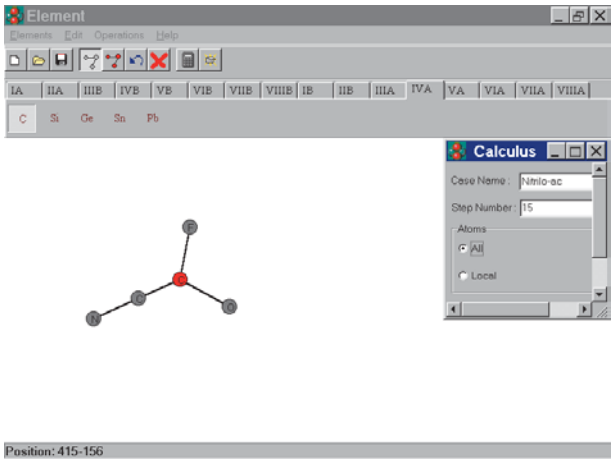


Fig. 1 Representation of nitrilo-acetyl fluoride in the graphical interface of MARCH-INSIDE

the electronegativity of the atoms that "compete" with a_j in order to withdraw electrons from a_i. The calculation of ${}^{SR}\pi_k$ for any organic or inorganic molecule was implemented in the software MARCH-INSIDE. [18] This software has a graphical interface to make it easier to use for medicinal chemists (see Fig. 1).

In Eq. (2) S_m represents a specific group of atoms in the molecule. When S_m contains all the atoms in the molecule, ${}^{SR}\pi_k(S)$ becomes a global molecular index and we write only ${}^{SR}\pi_k$. We can calculate different families of molecular descriptor by selecting different S_m conditions. For example, if we select only the halogen atoms, we write S_m=Halogens. Thence, we can use the following notation ${}^{SR}\pi_k$(Halogens) to represent the molecular descriptor calculated for this specific condition (S_m=Halogens). The same notation is used for other S_m, e.g., ${}^{SR}\pi_k$(Alq.) for S_m=Carbon atoms in aliphatic chains.

The 0th-step-self-return-electron-transition probabilities to core a_i (${}^0p_{ii}$) are the values of the main diagonal of ${}^0\pi=({}^1\Pi)^0=I_n$. I_n is the identity matrix of order n. Then ${}^0p_{ii}$ is, by definition, equal to 1 for any atom. Thence, ${}^{SR}\pi_0(S_m)$ is a count of the number of atoms that obey the condition S_m in the molecule. Therefore, ${}^{SR}\pi_0$=the total number of atoms, ${}^{SR}\pi_k$(Halogens)=the total number of halogens atoms and ${}^{SR}\pi_k$(Alq.)=the total number of carbon atoms in aliphatic chains in the molecule, respectively.

Statistical analysis

Following the previous section, we can try to develop a simple linear quantitative structure–activity relationship (QSAR) using the MARCH-INSIDE methodology with this general formula:

$$\text{Activity} = b + b_0D_0 + b_1D_1 + b_2D_2 + \ldots\ldots + b_kD_k \qquad (3)$$

Two LDAs [19, 20] were carried out. One model was used to discriminate fasciolicides from non-fasciolicide compounds. The other model was used to predict the mechanism of action of compounds. In both models the structure is represented by the molecular indices (${}^{SR}\pi_k$). In the first model the activity is coded by a dummy variable (Factv). This variable indicates either the presence (Factv=1) or absence (Factv=–1) of biological activity against *F. hepatica* or *F. gigantica*. In the second model, the activity is codified by a nominal variable (Mechanism). This variable is equal to 1 for compounds having as mechanism of action the inhibition of the β-tubulin function. [21, 22] The variable Mechanism is equal to 2 for such chemicals that are known as proton ionophores (see Martin references [21, 22]) or –1 for non-fasciolicide drugs. Chemicals that act against helminthes by a mechanism of action different from 1 and 2 but are not necessarily effective as fasciolicides were used in the predicting series as well as fasciolicides with different or unknown mechanism of action. In Eq. (3) b_k are the coefficients of the classification function, determined by the least squares method as implemented in the LDA modulus of STATISTICA 99'. [19] Forward stepwise was fixed as the strategy for variable selection. In both models we use the first 15 ${}^{SR}\pi_k$ in the MARCH-INSIDE and LDA model to develop the QSAR. We also use the local analogues of the above-mentioned molecular descriptors. [15] These local molecular descriptors were calculated selecting halogen atoms or aliphatic chains as explained above.

The quality of the model was determined by examining Wilk's λ statistic, Mahalanobis distance, Fisher ratio (F) and the corresponding p-level ($p(F)$) as well as the percentage of good classification and the proportion between the cases and variables in the equation. We also consider the linear discriminant canonical analysis statistics such as: canonical regression coefficient (R_{can}), chi-squared and its p-level ($p(\chi^2)$). [19, 20] Calculation of the percentages of good classification in the external prediction set permits us to carry out the validation of the models. Validation of the models was corroborated by means of leave-n-out cross validation procedures. [19, 20] Compounds in the external prediction set were never used to develop the classification function. Here we considered a general data set composed of 92 organic chemicals. The active compound series was built by joining together a group of 17 fasciolicide compounds that are currently in clinical use [3] with 27 nitrosalicylanilides extracted from an application patent [23] and eight compounds reported as fasciolicides. [24] The inactive compounds were selected from those considered as anthelmintics but not fasciolicides [21, 22] and another 21 structurally diverse chemicals reported by Côrba [24] as non-fasciolicides in primary screening tests.

All compounds were classified by a bracket-based system (a,b) (see Table 1) as follows: if a=+ it means that the compound has fasciolicide activity; a=– means a non-fasciolicide compound, b=t a compound that inhibits β-tubulin growth, b=H a compound that inhibits proton flow and b=0 a compound with an unknown mechanism of action. For example: the compounds that obey the

condition (+, 0) are fasciolicide drugs with an unknown mechanism of action.

Fasciolicide effectivity test

An in vivo experiment to measure the chemical effectiveness against *F. hepatica* was performed. An experimental technique reported in the literature [25] was selected for biological material processing and *F. hepatica* egg extraction. Mitterpak et al.'s technique for host (*Lymnaea cubensis*) invasion was carried out. [26] Afterwards we followed the steps reported by Olazábal et al. [27] to obtain the metacercariae. Metacercariae were conserved in the cold until the in vivo experiment. [25]

Balb/c mice were selected as the biological model. Healthy Balb/c mice of both sexes and food were purchased from the "Centro Nacional de Animales de Laboratorio (CENPALAB)", Cuba. Quarantine, labeling, acclimatization and good maintenance conditions of animals were strictly obeyed. [25, 28] The CBQ organic synthesis laboratory synthesized the compounds A–F, with 98% purity. These chemicals were tested in order to evaluate their effectiveness against *F. hepatica* according to the following experimental design. Eight treatment groups with five mice per group were created. One group (infected control group) was treated with sunflower oil (administration vehicle). The second group was neither infested nor treated. Each mouse in the six remaining groups was treated with one of the compounds (A–F). The compounds were previously diluted in 10 ml of sunflower oil to obtain a single dose of 200 mg per kg of body weight. The solutions were used immediately after preparation. All the products were administered by the oral route. All mice received 0.2 ml of each solution with a syringe of 1 ml. Mouse invasion with metacercariae of *F. hepatica*, 2 weeks old, 14 days before drug administration, was carried out following Côrba et al.'s methodology. [24]

The effectiveness was measured based on the elimination or not of *Fasciola hepatica*, in their juvenile stage, as shown by laboratory diagnostics. From different effectiveness indexes [29, 30, 31] the *E%* index was selected. This is a quantitative indicator of effectiveness introduced by Steward [32] and defined as $E\%=[(XC–XT)/XC]\times100$. Here *E%* is the percentage of effectiveness, XC is the average amount of *Fasciola* in the control group and XT is the average amount of *Fasciola* in the treated group. We used this index in spite of the existence of other (more recently defined) effectiveness parameters because it is a direct expression of effectiveness that we can compare with $\Delta P\%$.

Results

Once we perform a random and representative selection of training set, it can be used to fit the discriminant function. The model selection was subjected to the principle of parsimony. Thus, we chose the functions with higher statistical signification but with as few parameters (b_k) as possible:

MARCH-INSIDE and LDA fasciolicide activity classification function:

$$\begin{aligned}\text{Factv} &= 2.0281^{\text{SR}}\pi_0\,(\text{Halogens}) - 0.7926^{\text{SR}}\pi_0\,(\text{Alq.}) \\ &\quad + 1.9056^{\text{SR}}\pi_{12} - 8.7961 \\ N = 65\lambda &= 0.36F\,(3,61) = 36.3D^2 = 7.15p\,(F) \\ &< 0.00p\,(\chi^2) < 0.00R_{\text{can}} = 0.80\end{aligned} \tag{4}$$

MARCH-INSIDE and LDA fasciolicide mechanism of action classification function:

$$\begin{aligned}\beta\text{–Tubulin}\,(1) &= 1.328^{\text{SR}}\pi_0\,(\text{Halogens}) \\ &\quad - 1.192^{\text{SR}}\pi_0\,(\text{Alq.}) - 262.609^{\text{SR}}\pi_{11} \\ &\quad + 2.548^{\text{SR}}\pi_0 + 277.886^{\text{SR}}\pi_{13} \\ &\quad - 3.477^{\text{SR}}\pi_1 - 13.918\end{aligned} \tag{5}$$

$$\begin{aligned}\text{H} - \text{Ionophores}\,(2) &= 2.714^{\text{SR}}\pi_0\,(\text{Halogens}) \\ &\quad - 0.936^{\text{SR}}\pi_0\,(\text{Alq.}) + 147.941^{\text{SR}}\pi_{11} \\ &\quad - 3.599^{\text{SR}}\pi_0 - 137.679^{\text{SR}}\pi_{13} \\ &\quad + 6.206^{\text{SR}}\pi_1 - 22.477\end{aligned} \tag{6}$$

$$\begin{aligned}\text{Non} - \text{active}\,(-1) &= 1.042^{\text{SR}}\pi_0\,(\text{Halogens}) \\ &\quad - 0.671^{\text{SR}}\pi_0\,(\text{Alq.}) - 173.561^{\text{SR}}\pi_{11} \\ &\quad + 2.706^{\text{SR}}\pi_0 + 184.277^{\text{SR}}\pi_{13} \\ &\quad - 5.265^{\text{SR}}\pi_1 - 11.305 \\ N = 65\lambda &= 0.09F\,(12,116) = 22.25p\,(F) < 0.00\chi^2 \\ &= 144.5p\,(\chi^2) < 0.00R_{\text{can}} = 0.945\end{aligned} \tag{7}$$

Here λ is the Wilks statistic, which, for overall discrimination, takes values in the range from 0 (perfect discrimination) to 1 (no discrimination). The Fisher test permits us to test the hypothesis of separation of groups with a probability of error (*p*-level) $p(F)<0.05$.

The first model correctly classified 93.85% of the compounds in the training series, i.e., four misclassifications in 65 cases, while in the predicting set there were two errors in 19 cases, that is, 89.5% good classification. Specifically, the model classified 97.06% of fasciolicide compounds in the training set correctly and 80% of these compounds in the predicting set, i.e., two misclassifications in ten cases. On the other hand, the model correctly classified 90.32% of non-fasciolicide compounds in the training set and 100% of these compounds in the predicting set. The names or code of all compounds used to derive the QSAR as well as their predicted activity by both models are shown in Table 2.

On the other hand, the results obtained after evaluation of external predicting sets with both models are depicted in Table 3. The chemical substituents of the encoded compounds appear in Table 4 and the respective molecular skeletons are shown in Fig. 2.

In Tables 2 and 3 $\Delta P\%=[P(\text{actv})–P(\text{non-actv})]\times100$, where P(actv) is the probability that the equation classi-

Table 2 Training set classification results

Training set

Name or code[a]	ΔP%[b]	P(–)[c]	P(t)[c]	P(H)[c]	Name or code[a]	ΔP%[b]	P(–)[c]	P(t)[c]	P(H)[c]
[f]Albendazole (+,t)[d]	–98.8	32.0	68.0	0.0	Hexacloroethane (+,0)	99.3	1.5	97.8	0.7
Bitionol (+,0)	97.8	0.0	0.0	100.0	Meniclofolan (+,H)	58.7	0.0	0.0	100.0
Brotianianide (+,H)	94.9	0.0	0.0	100.0	Oxyclozanide (+,H)	99.8	0.0	0.0	100.0
Closantel (+,H)	99.9	0.0	0.0	100.0	Rafoxanide (+,H)	99.6	0.0	0.0	100.0
Dibromsasalam (+,H)	13.4	23.5	10.5	66.0	Tetra-Cl-ethane (+,0)	99.3	1.4	83.6	15.1
Dioxapiramizole(+,0)	99.0	29.1	70.9	0.0	Carbontetrachloride(+,0)	54.5	0.3	99.7	0.0
Hexaclorfeno (+,H)	99.6	0.0	0.0	100.0	Metronidazole (–)	–99.8	69.0	30.9	0.0
6-Cloroquine (–)	–95.9	96.4	3.6	0.0	Netobimin (–)	–79.8	5.9	94.1	0.0
Butamizole (–)	–72.3	38.1	61.9	0.0	Oxantel (–)	–93.3	68.2	31.8	0.0
Cambendazole (+,t)	–70.4	24.2	75.8	0.0	Oxfendazol (–,t)	–85.3	37.1	62.9	0.0
Diaveridine (–)	–97.8	79.5	20.5	0.0	Pyrantel (–)	–99.7	53.0	47.0	0.0
DiEt-carbazide (–)	–99.8	92.6	7.4	0.0	Praziquantel (–)	–99.5	99.6	0.4	0.0
Fenotiazine (–)	–98.8	69.9	30.1	0.0	Thiabendazol (–,t)	–98.7	24.8	75.2	0.0
Mebendazole (–)	–72.1	61.2	38.8	0.0	11754 (–)	–93.2	35.9	64.1	0.0
Methyridine (–)	–97.6	23.7	76.3	0.0	11755 (–)	–99.6	88.1	11.9	0.0
10448 (–)	–81.4	60.6	39.4	0.0	11758 (–)	–98.0	48.5	51.5	0.0
10451 (–)	–84.0	65.5	34.5	0.0	11780 (–)	–100.0	99.6	0.4	0.0
[f]10477 (–)	10.8	21.4	78.6	0.0	11902 (–)	–23.3	38.6	61.4	0.0
11463 (–)	–48.6	19.6	80.4	0.0	8217 (–)	–35.1	11.3	88.7	0.0
11561 (–)	–81.2	91.5	8.4	0.0	8218 (–)	–99.0	83.1	16.9	0.0
[f]11562 (–)	60.9	72.2	27.8	0.0	9298 (–)	–99.8	88.0	12.0	0.0
[f]11564 (–)	77.8	61.1	38.8	0.1	11567 (–)	–93.1	90.3	9.7	0.0

R1[e]	R2[e]	ΔP%[b]	P(–)[c]	P(t)[c]	P(H)[c]	R1[e]	R2[e]	ΔP%[b]	P(–)[c]	P(t)[c]	P(H)[c]
F	F	96.0	0.0	0.0	100.0	Br	F	100.0	0.0	0.0	100.0
F	Cl	97.7	0.0	0.0	100.0	Br	Cl	87.1	0.0	0.0	100.0
CH_3	Br	95.7	0.0	0.0	100.0	CH_3	Br	73.2	0.0	0.0	100.0
F	I	95.6	0.0	0.0	100.0	Br	I	72.9	0.0	0.0	100.0
CF_3	CF_3	98.1	0.0	0.0	100.0	Br	CF_3	99.7	0.0	0.0	100.0
F	CH_3	88.1	0.0	0.0	100.0	CH_3	CH_3	83.7	0.0	0.0	100.0
Cl	F	87.5	0.0	0.0	100.0	I	F	83.7	0.0	0.0	100.0
Cl	Cl	73.7	0.0	0.0	100.0	I	Cl	74.7	0.0	0.0	100.0
CF_3	Br	71.6	0.0	0.0	100.0	CH_3	Br	90.6	0.0	0.0	100.0
Cl	I	70.6	0.0	0.0	100.0	I	I	73.3	0.0	0.0	100.0
Cl	CF_3	88.7	0.0	0.0	100.0	I	CF_3				

[a] The chemical substituents of the drugs represented here by a code are depicted in Table 2 and the respective molecular core in the Fig. 1
[b] Fluckicidal activity predicted by model 1 (see material and methods, statistical analysis)
[c] Percentage of probability with which the drug is predicted as non-fluckicidal, having the β-tubulin or proton ionophore mechanisms of action, respectively
[d] For explanation of the brackets notation see the materials and method section
[e] Chemical substituents of the drug with reference to the basic framework of nitrosalicyl anilides (see Fig. 2)
[f] Compounds that are misclassified by the fasciolicide/non-fasciolicide discriminant function

Table 3 Predicting set classification results

Predicting set

Name or codea	ΔP%[b]	P(–)[c]	P(t)[c]	P(H)[c]	Name or codea	ΔP%[b]	P(–)[c]	P(t)[c]	P(H)[c]
Clorsulon (+,0)[d]	93.5	0.0	0.0	100.0	[f]Nitroxinyl (+,H)	–90.9	0.0	0.0	100.0
[f]Disofenol (+,0)	–61.0	11.8	7.3	80.8	Triclabendazole (+,t)	88.4	27.7	72.2	0.2
D.F.D. (–)	–64.5	60.5	39.5	0.0	Morantel (–)	–99.3	72.1	27.9	0.0
Levamizole (–)	–97.0	31.3	68.7	0.0	Pentamidne (–)	–92.5	94.3	5.4	0.3
11 532 (+)	94.6	40.8	57.8	1.3	10 449 (–)	–99.4	96.2	3.8	0.0
11 534 (+)	85.2	0.0	0.0	100.0	11 542 (–)	–73.4	21.3	78.7	0.0
11 757 (–)	–93.0	47.6	52.4	0.0	11 566 (–)	–30.3	92.9	7.1	0.0
8 216 (–)	–52.4	19.4	80.6	0.0					

R1[e]	R2[e]	ΔP%[b]	P(–)[c]	P(t)[c]	P(H)[c]	R1[e]	R2[e]	ΔP%[b]	P(–)[c]	P(t)[c]	P(H)[c]
F	I	95.6	0.0	0.0	100.0	CL	Br	100.0	0.0	0.0	100.0
CF_3	F	88.3	0.0	0.0	100.0	Br	I	99.5	0.0	0.0	100.0
CH_3	CF_3	71.2	0.0	0.0	100.0	I	CF_3	74.6	0.0	0.0	100.0

[a] The chemical substituents of the drugs represented here by a code are depicted in Table 2 and the respective molecular core in the Fig. 1
[b] Fluckicidal activity predicted by model 1 (see material and methods, statistical analysis)
[c] Percentage of probability with which the drug is predicted as non-fluckicidal or having the β-tubulin or a proton ionophore mechanisms of action, respectively; using Eqs. (5), (6) and (7)
[d] For explanation of the brackets notation see the materials and method section
[e] Chemical substituents of the drug with reference to the basic framework of nitrosalicyl anilides (see Fig. 2)
[f] Compounds that are misclassified by the fasciolicide/non-fasciolicide discriminant function

Table 4 Chemical substituents of drugs represented in Table 1 by a code

Chinazolinones

Code[a]	R
10 448	Phenylamine
10 449	Cyclohexylamine
10 451	Morpholine

Chinoxalines

Code[a]	R	X
8 216	H	Cl
8 217	CH3	Cl
8 218	$H_3C–(CH_2)_3–$	Cl
9 298	H3C–(CH2)11–	NH2

Piperazines (I)

Code[a]	R	Code[a]	R
11 534	H	11 562	$H_3C–(CH_2)_2–$
11 532	CH_3	11 566	$H_3C–CH_2–$
11 567	$H_3C–(CH_2)_5–$		

Piperazines (II)

Code[a]	R
11 561	$H_3C–(CH_2)_2–$
11 567	$H_3C–(CH_2)_5–$

Pyrimidines (I)

Code[a]	R_1	R_2
11 758	H	H
11 780	H	Cyclooctyl

Pyrimidines (II)

Code[a]	X
11 755	CH_2
11 754	O
11 757	S

Benzimidazolurethans

Code[a]	R
11 902	(CH3)2N
11 542	H

Code[a] R
11 463 –OCH3
10 477 Cl

[a] This code is the same as reported by Côrba [24]

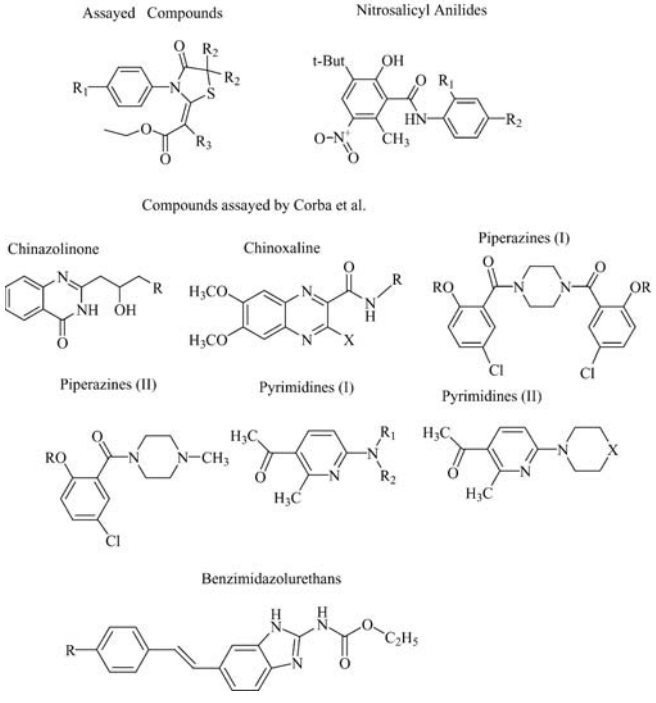


Fig. 2 Basic molecular skeleton of compounds used to develop the QSAR

fies a compound as active. Conversely, *P*(non-actv) is the probability that the model classifies a compound as non-active. This value ($\Delta P\%$) takes positive values when *P*(actv)>*P*(non-actv) and negative otherwise. Therefore, when $\Delta P\%$ is positive (negative) the compound was classified as fluckicidal (non-fluckicidal). When $\Delta P\%$ is in the range $-5<\Delta P\%<5$ the compound was considered as unclassified. [15, 19]

For a more exhaustive testing of the predictive power of the model, we carried out leave-*n*-out cross validation procedures. These validation techniques are implemented in the module for classification trees training in Statistica 99'. [19] In this module, the user can select discriminant-based linear combination splits as the split method, prune on misclassification error as the stopping rule and the same prior probabilities as in Eq. (4) and obtains this equation as the split rule. Once Eq. (4) is modeled in the classification trees' module the folding parameter of the cross validation can be varied to carry out the leave-*n*-out procedure. This model shows 87.7, 93.8, 92.2 and 93.9% of global good classification when *n* varies from 2 to 6 in leave-*n*-by time cross validation procedures. The model seems to be stabilized at around 92% of good classification when *n* is >6 (see Fig. 3).

The second model (we report three equations because this is a three group classification problem) classifies 70% of non-fasciolicide compounds, 85.71% of β-tubulin inhibitors and 100% of proton ionophores in the training set. This model recognizes as proton ionophores 100% of any nitrosalicylanilides in the predicting series. We show the graphical results of canonical analysis in Fig. 4.

The results of the efficacy in preliminary screening and the probabilities predicted by the model for six experimentally tested compounds supplied by the Chemicals Bio-active Center are shown in Table 5.

Discussion

The resistance of flukes *F. hepatica* and *F. gigantica* to fasciolicides has begun. [8, 3] For example, Moll et al. reported (in the year 2000) resistance of fasciolosis to triclabendazole, which is usually an effective treatment. [33] However, we cannot wait until the fasciolicide-resistance problem becomes uncontrollable to begin the development of novel methods for fasciolicide drug selection. The discriminant function developed here is

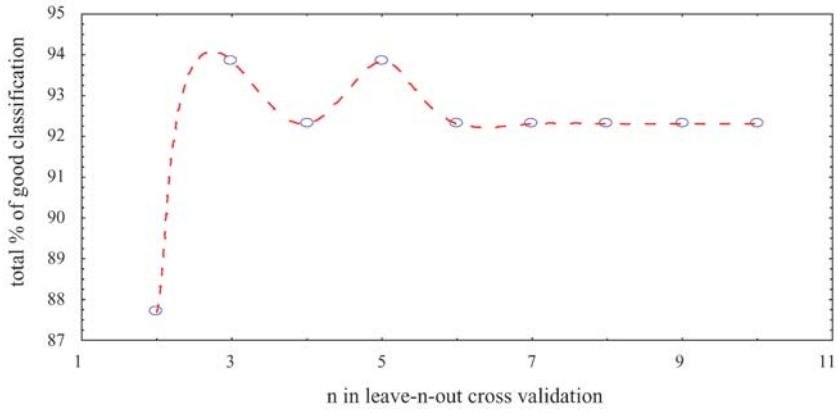
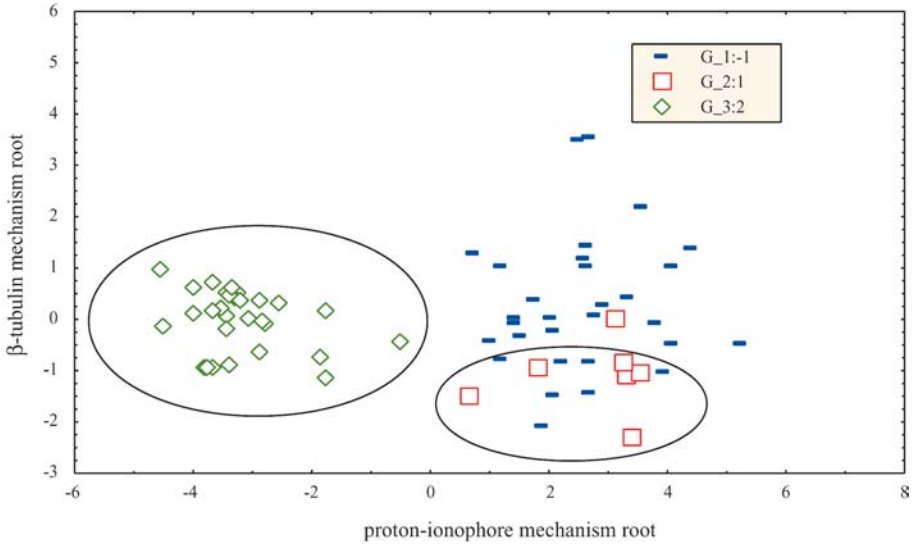


Fig. 3 Behavior of the global or total percentage of good classification in different *n*-fold cross-validation analysis



G1:-1 is the group of non-active compounds,
G2:1 is the group of β-tubulin inhibitors,
G3:3 is the group of proton-ionophores.

Fig. 4 Graphical representation of the results of the canonical analysis

Table 5 Comparison between experimental and theoretical activities

Code[a]	Structure[b]	*E*%[c]	Δ*P*%[d]	*P*(–)[e]	*P*(t)[e]	*P*(H)[e]
A	R1=p-OCH_3Phenyl, R2=Cl, R3=CN	88.4	88.6	20.2	79.8	0.0
B	R1=Phenyl, R2=Cl, R3=CN	88.4	75.9	24.3	75.7	0.0
C	R1=Phenyl, R2=H, R3=SCH_3	0	15.6	12.8	87.2	0.0
D	R1=p-OCH_3Phenyl, R2=H, R3=SCH_3	0	–28.3	2.1	97.9	0.0
E	R1=Furan, R2=H, R3=SCH_3	0	–94.6	24.4	75.6	0.0
F	R1=p-CH_3Phenyl, R2=H, R3=SCH_3	0	–4.0	13.5	86.5	0.0

[a] This code is used by CBQ
[b] Chemical substituents of the drug refer to the basic framework of the compounds experimentally assayed here (see Fig. 2)
[c] Biological activity as measured in the experimental assay
[d] Posterior probability of activity predicted with Eq. (4)
[e] Percentage of probability with which the drug is predicted to be non-fluckicidal or having a β-tubulin or a proton ionophore mechanism of action, respectively using Eqs. (5), (6) and (7)

based on a single molecular modeling related methodology (MARCH-INSIDE) and facilitates the solution of this problem. This model did not misclassify any non-fasciolicide compound in the prediction series and shows an acceptably low 9.38% of false actives in the training set. This means that, if we use this model as a guide for compound selection in a screening program, we can expect (with 95% probability) that only nine of 100 chemi-

cals submitted to experimental assays will be an ineffective fasciolicide. There is a major interest in ensuring the minimization of false active compound selection because it causes a great loss of time and resources. [12] Classification Eq. (4) was fitted using 65 cases (compounds). Therefore, the seven-fold cross validation procedure is, approximately, a leave-10%-out cross validation procedure (6.5 cases are 10% of 65 cases). However, as depicted in Fig. 3, Eq. (4) has a very stable predictive behavior when *n* is larger than 6 in leave-*n*-out cross validation procedures. All the leave-*n*-out percentages of good classification are higher than 85%. This value is considered as a threshold limit to accept a model as valid. [9, 13]

A direct inspection of the QSAR reported here shows us that the number of halogen atoms in the molecule ($^{SR}\pi_0$(Halogen)) increases approximately 2.03 times the probability that a compound acts as a fasciolicide in spite of the mechanism of action. This can be explained if we consider that almost all fasciolicide chemicals studied here act either as β-tubulin inhibitors or proton ionophores. Both mechanism of action require the existence of electron-withdrawing groups that facilitate either the dissociation of weak acid groups (such as phenol and amide) or the activation of electrophilic centers in the molecule. [21, 22, 34, 35, 36, 37] The second LDA also agrees with this explanation. On the other hand, the number of aliphatic chains decreases the biological activity. This can be explained because these structural features facilitate the distribution of the drug to lipoid tissues, decreasing the possibility of interaction with the pharmacological target in the cell. [38]

There are some interesting cases that we going to discuss here. For example, albendazole was misclassified by Eq. (4) but it was predicted very well to be a β-tubulin mechanism acting drug by the second model. This result suggests that results of the two models must be matched before arriving at any conclusions about the activity of a compound.

On the other hand, the mode of action of simple halogenated derivatives against *F. hepatica* is not fully understood. [39] Hexachloroethane, carbon tetrachloride and tetrachloroethane are recognized as fluckicidal drugs but predicted as β-tubulin inhibitors. This is a logical result if we consider the great activation that halogen atoms cause in the electrophilic reactive centers of these molecules. [40] Nevertheless, this result may be considered as preliminary and must be subjected to experimental corroboration.

The canonical analysis (Fig. 4) detected with a significant regression coefficient (Rc=0.945, $p<0.05$) the presence of three clusters of compounds. The existence of non-active chemicals (against flukes) that could act by the β-tubulin mechanism against another helminthes is an experimental fact that could cause the overlap between β-tubulin inhibitors and the non-fluckicidal drug clusters detected in the canonical analysis (see Fig. 4). Thus, we can conclude that having a mechanism of action against helminthes does not exclude the drug from acting by another mechanism and is not a specific condition to act specifically against flukes.

Finally, very good agreement between the experimental fasciolicide activity test and the predicted activities for six compounds was found. As depicted in Table 5, compounds A and B are predicted as fluckicidal compounds and have an effectivity greater than 80%. The remaining compounds did not show any fluckicidal power in the biological assay and were predicted as non-active compounds. Thus, the model shows an overall 100% efficacy in the experimental assessment.

Conclusions

Virtual screening has emerged as an interesting alternative to high-throughput screening. [12] Thus, the continuous definition of novel molecular descriptors that could explain different pharmacological properties by means of a QSAR is necessary. [41, 42, 43] Consequently, we have developed two MARCH-INSIDE and LDA models that could permit us to predict by fast in silico screening the fasciolicide activity of chemicals and to outline preliminary conclusions about possible mechanisms of action.

Acknowledgements The authors would like to express their gratitude to Ph.D. Jose Luis Garcia and the Cuban Ministry of Higher Education for financial support. González DH thanks the University of "Santiago de Compostela" in Spain for kind hospitality and the Xunta the Galicia (PR411A 2001/48-0) for partial financial support.

References

1. Shah-Fischer M, Say RR (1989) Manual of tropical veterinary parasitology. CAB International, Wallingford, UK, pp 63–72
2. Benenson A (1997) El control de las enfermedades trasmisibles en el hombre. OPS, No 507
3. Fairweather I, Borav J (1999) Vet J 158:81–112
4. Hillyer G, Apt W (1997) Parasitol Today 13:87–88
5. Anon (1995) WHO Technical series. Guevara, 849, p 157
6. O'Neill S, Parkinson M, Strauss W (1998) Am J Trop Med Hygiene 58:417–428
7. More A, Castilo N, Borges M, Fernández N, Osonio M (1997) Fascioliasis humana: manifestaciones clínicas y diagnóstico en pacientes en el municipio de Santa Clara. XIII Congreso Latinoamericano de Parasitología, La Habana, Cuba, p 126
8. Boray J, De Bono D (1989) Austr Vet J 166–169
9. Gálvez J, García R (1994) Diseño de fármacos por conectividad molecular. In: Mosquiera A (ed) Diseño de Medicamentos. Farmaindustria, Madrid, pp 355–384
10. Kubinyi H, Taylor J, Ramdsen C (1990) Quantitative drug design. In: Hansch C (ed) Comprehensive medicinal chemistry, vol 4. Pergamon, Oxford, pp 589–643
11. Lajiness M (1990) Molecular similarity-based methods for selecting compounds for screening. In: Rouvray DH (ed) Computacional chemical graph theory. Nova Science, New York
12. Estrada E, Uriarte E, Montero A, Teijeira M, Santana L, De Clercq E (2000) J Med Chem 10:1975–1985
13. Estrada E, Peña A, García-Domenech R (1998) J Comput-Aided Mol Des 12:583–595
14. Todeschini R, Consonni V (2000) Handbook of molecular descriptors. Wiley VCH, Weinheim, Germany

15. González H, Hernádez I, Ramos R, Cruz M, Akpaloo D, Torres G, Cabrera M (2002) J Mol Model (submitted)
16. Gnedenko B (1978) The theory of probability. Mir, Moscow, pp 107–112
17. Pauling L (1939) The nature of chemical bond. Cornell University Press, Ithaca, N.Y. pp 2–60
18. Hernández I, González DH (2002) **MARCH-INSIDE** version 1.0 (**Mar**kovian **che**micals "**in si***lico*" **de**sign). Chemicals Bioactives Center, Central University of "Las Villas", Cuba. This is a preliminary experimental version; in future a professional version will be available to the public. For any information about it send an e-mail to the corresponding author humbertogd@navegalia.com.
19. STATISTICA 99', version 5.5 (1999) Statsoft Inc
20. Johnson RA, Wichern DW (1988) Applied multivariate statistical analysis. Prentice-Hall, Englewood Cliffs, N.J.
21. Martin R (1997) Vet J 154:11–34
22. Martin R, Robertson A, Bjorn H (1997) Parasitology 114:S111-S124
23. Meek W (1975) USP, appl 465:884
24. Čorba J, Velebny S, Spaldonova R (1981) Helminthologia 18:43–52
25. Milligen FJ, Cornelissen JB, Guasenbeek CP, Bokhout BA (1998) J Immunol Methods 213:183–90
26. Mitterpak J, Mendéz M, Mauri M (1972) Serie biológica 30:1–24
27. Olazábal E, Morales A, Serrano H, Brito E (1999) Vet Méx 30:109–114
28. Ping C, Hayes A (1994) Acute toxicity and eyes irritancy. In: Wallace A (ed) Principles and methods of toxicology, 3rd edn. Hayes, Raven Press, New York
29. Tsocheva N, Krustev L, Polyakova O (1992) Helminthologia 27:261–268
30. Poitou I, Baeza E, Boulard C (1992) Vet Parasitol 45:59–71
31. Espaine L, Lines R, Demedio J (1996) Manual de Parasitología y enfermedades Parasitarias I. Editorial MES, Habana, p 103
32. Steward JS (1955) Parasitology 45:231–241
33. Moll A, Gaasenbeek CPH, Vellema P, Borgsteede FHM (2000) Vet Parasitol 91:153–158
34. Mckellar QA, Kinabo LDB (1991) Br Vet J 147:306–319
35. Stryer L (1995) Biochemestry. WH Freeman, New York
36. Lacey (1990) Parasitol Today 6:112–115
37. Alvinerie M, Floch R, Galter P (1995) J Vet Pharmacol Therapuetic 14:170–173
38. Gibaldi M, Perrier D (1982) Pharmacokinetics. Marcel Dekker, New York, pp 145–265
39. Mc Kellar AQ, Kinabo DBL (1991) Br Vet J 147:306–318
40. Szabo A, Ostlund NS (1989) Modern quantum chemistry. McGraw-Hill, New York
41. Öberg T (2001) Internet J Chem 4:7
42. Mansfield ML, Covell DG, Jernigan RL (2002) J Chem Inf Comput Sci 2:259–273
43. Dury L, Latour T, Leherte L, Barberis F, Vercauteren DP (2001) J Chem Inf Comput Sci 6:1437–1445