Available online at www.sciencedirect.com
SCIENCE DIRECT®
Bioorganic & Medicinal Chemistry 13 (2005) 1005–1020
Bioorganic & Medicinal Chemistry

# Atom, atom-type and total molecular linear indices as a promising approach for bioorganic and medicinal chemistry: theoretical and experimental assessment of a novel method for virtual screening and rational design of new lead anthelmintic

Yovani Marrero-Ponce,[a,b,*] Juan A. Castillo-Garit,[b,c] Ervelio Olazabal,[d] Hector S. Serrano,[d] Alcidez Morales,[d] Nilo Castañedo,[b,d] Froylán Ibarra-Velarde,[e] Alma Huesca-Guillen,[e] Alicia M. Sánchez,[a] Francisco Torrens[f] and Eduardo A. Castro[g]

[a]*Department of Pharmacy, Faculty of Chemical-Pharmacy, Central University of Las Villas, Santa Clara, 54830 Villa Clara, Cuba*
[b]*Department of Drug Design, Chemical Bioactive Center, Central University of Las Villas, Santa Clara, 54830 Villa Clara, Cuba*
[c]*Applied Chemistry Research Center, Central University of Las Villas, Santa Clara, 54830 Villa Clara, Cuba*
[d]*Department of Parasitology, Chemical Bioactive Center, Central University of Las Villas, Santa Clara, 54830 Villa Clara, Cuba*
[e]*Department of Parasitology, Faculty of Veterinarian Medicinal and Zootecnic, UNAM, Mexico, D.F. 04510, Mexico*
[f]*Institut Universitari de Ciència Molecular, Universitat de València, Dr. Moliner 50, E-46100 Burjassot, València, Spain*
[g]*INIFTA, División Química Teórica, Suc.4, C.C. 16, La Plata 1900, Argentina*

Received 6 October 2004; revised 17 November 2004; accepted 22 November 2004
Available online 16 December 2004

**Abstract**—Helminth infections are a medical problem in the world nowadays. In this paper a novel atom-level chemical descriptor has been applied to estimate the anthelmintic activity. Total and local linear indices and linear discriminant analysis were used to obtain a quantitative model that discriminates between anthelmintic and non-anthelmintic drug-like compounds. The discriminant model has an accuracy of 90.11% in the training set, with a high Matthews' correlation coefficient (MCC = 0.80). To assess the robustness and predictive power of the obtained model, internal (leave-*n*-out) and external validation process was performed. The QSAR model correctly classified 88.55% of compounds in this external prediction set, yielding a MCC of 0.77. Another LDA model was carried out to outline some conclusions about the possible modes of action of anthelmintic drugs. It has an accuracy of 93.50% in the training set, and 80.00% in the external prediction set. After that, the developed model was used in the virtual—in silico—screening and several compounds from the Merck Index, Negwer's Handbook and Goodman and Gilman were identified by the model as anthelmintic. Finally, the experimental assay of an organic chemical (a furylethylene derivative) by an in vivo test permits us to carry out an assessment of the model. An accuracy of 100% with the theoretical predictions was observed. These results suggest that the proposed method will be a good tool for studying the biological properties of drug candidates during the early state of the drug-development process.


## 1. Introduction

Anthelmintic drugs are used to control, prevent, and treat nematode, cestode, and trematode parasite infestations in humans and domestic animals.[1] Regrettably, with the increased use of these compounds, anthelmintic resistance (see Table 1) has appeared and increased in frequency.[2] If resistance to a particular anthelmintic has occurred, it is likely that another anthelmintic with the same mode of action will also be ineffective although other anthelmintics with another mode of action.[3] Moreover, continued economic losses in animal production and human diseases due to parasites are still of concern to industrial chemists looking for new anthelmintic agents.[1] The great cost associated with the development

*Keywords*: Anthelmintic activity; QSAR; TOMOCOMD-CARDD software; Total and local linear indices; Virtual screening.
* Corresponding author. Tel.: +53 42 281192/281473; fax: +53 42 281130/281455; e-mail addresses: yovanimp@qf.uclv.edu.cu; ymarrero77@yahoo.es

0968-0896/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmc.2004.11.040

**Table 1.** Resistances to anthelmintics

| Host | Parasite | Drug |
|---|---|---|
| Humans | Schistosomes[5,6] | Oxamniquine<br>Praziquantel |
| | Hookworms[7,8] | Tetrahydropyrimidines |
| Extensive grazing pigs | Trichostrongylids[9,10] | Benzimidazoles |
| | | Macrocyclic lactones |
| | Oesophagostomum[9,10] | Benzimidazoles |
| | | Macrocyclic lactones<br>Tetrahydropyrimidines |
| Sheep and goats | Trichostrongylids[9,10] | Benzimidazoles |
| | | Levamisole<br>Tetrahydropyrimidines<br>Macrocyclic lactones |
| | Fasciola[9,11,12] | Benzimidazoles<br>Triclabendazol<br>Closantel |
| Horses | Small strongyles[9,10] | Benzimidazoles<br>Tetrahydropyrimidines<br>Piperazine |

of new compounds and the small market for anthelmintics makes slow this development.[4]

At present, it is not possible to carry out a rigorous classification of the anthelmintic drugs according to their mechanisms of action. However, there are two major modes of action of these drugs: (i) the first one group acts on the ion channels of parasite membranes and usually has more rapid therapeutic effect, and (ii) the second group acts more slowly on a range of biochemical target sites found in parasites (see original work in Ref. 1).

At present, virtual (computational) screening[13] of chemical libraries has emerged as a complementary approach to high-throughput screening (HTS).[14,15] By this means, computational techniques are used to select a reduced number of potentially active compounds from large available chemical or combinatorial libraries. This in silico procedure will be used here in order to find predictive models that permit us the 'rational' selection/identification or design of new anthelmintics with the required properties.[16,17]

Recently, a novel scheme to the rational—in silico—molecular design (or selection/identification of chemicals) and to QSAR/QSPR studies has been introduced by one of the present authors. It is the so-called Topological MOlecular COMputer Design (TOMOCOMD).[18] This method has been developed to generate molecular descriptors based on the linear algebra theory. In this sense, atom, atom-type, and total quadratic and linear indices have been defined in analogy to the quadratic and linear mathematical maps.[19,20] This approach has been successfully employed in QSPR[19,21] and QSAR[20,22] studies, including studies related to nucleic acid–drug interactions.[23] The approach describes changes in the electron distribution with time throughout the molecular backbone. The TOMOCOMD-CARDD (acronym of the Computed-Aided 'Rational' Drug Design) strategy is very useful for the selection of novel subsystems of compounds having a desired property/activity, which can be further optimized by using some of the many molecular modeling methods at the disposition of the medicinal chemists. The method has also demonstrated flexibility in relation to many different problems. In this sense, the TOMOCOMD-CARDD approach has been applied to the fast-track experimental discovery of novel antimalarial compounds.[25] The prediction of the physical, chem-physical, and chemical properties of organic compounds is a problem that can also be addressed using this approach.[19,26] Codification of chirality and other 3D structural features constitute another advantage of this method.[27] The latter opportunity has allowed the description of the significance-interpretation and the comparison to other molecular descriptors.[20,26]

The article is organized as follows. First, the TOMOCOMD-CARDD method (atom, atom-type, and total linear indices) is used to find a quantitative model that discriminates anthelmintic compounds from the inactive ones. Next, these total and local molecular descriptors are used in the generation of a discriminant function that permits the classification of anthelmintic compounds taking into account the different modes of action. After that, is performed a simulated experiment of virtual—in silico—screening for the search of new lead. Finally, an experimental assay (by an in vivo test) of a 2-furylethylene derivative is carried out, with the aim to prove the predictive power of the obtained models.

## 2. Theoretical approach: atom, atom-type, and total linear indices

The atom, atom-type, and total linear indices of the 'molecular pseudograph's atom adjacency matrix' for small-to-medium sized organic compounds have been explained in some detail elsewhere.[20] However, an overview of this approach will be given.

For a given molecule composed of $n$ atoms, the 'molecular vector' ($\boldsymbol{X}$) is constructed and the $k$th atom linear indices, $\boldsymbol{f}_k(x_i)$ are calculated as a linear maps on $\Re^n[\boldsymbol{f}_k(x_i) : \Re^n \rightarrow \Re^n$; thus $\boldsymbol{f}_k(x_i)$: endomorphism on $\Re^n]$ in canonical basis as shown in Eq. 1,

$$\boldsymbol{f}_k(x_i) = \sum_{j=1}^{n} {}^k a_{ij} X_j \tag{1}$$

where ${}^k a_{ij} = {}^k a_{ji}$ (symmetric square matrix), $n$ is the number of atoms of the molecule, and $X_1, \ldots, X_n$ are the coordinates or components of the 'molecular vector' ($\boldsymbol{X}$) in a system of canonical basis vectors of $\Re^n$. The components of the 'molecular' vector are numeric values, which can be considered as weights (atom-labels) for the vertices of the pseudograph. Certain atomic properties (electronegativity, density, atomic radius, etc.) can be used with this propose. In this work Mulliken electronegativity was selected as atom weights.[28]

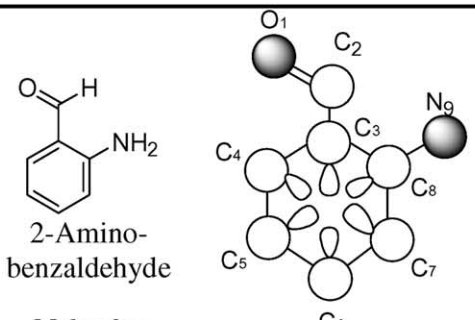
The coefficients ${}^{k}a_{ij}$ are the elements of the $k$th power of the symmetric square matrix **M**(G) of the molecular pseudograph (G) and are defined as follows:[19–27]

$$\begin{aligned} a_{ij} &= P_{ij} \quad \text{if } i \neq j \quad \text{and} \quad \exists e_k \in \mathrm{E(G)} \\ &= L_{ii} \quad \text{if } i = j \\ &= 0 \quad \text{otherwise} \end{aligned} \tag{2}$$

where E(G) represents the set of edges of G. $P_{ij}$ is the number of edges (bonds) between vertices (atoms) $v_i$ and $v_j$ and $L_{ii}$ is the number of loops in $v_i$ (see Table 2).

Note that atom linear indices are defined as a linear transformation $f_k(x_i)$ on a molecular vector space $\Re^n$. This map is a correspondence that assigns to every vector $X$ in $\Re^n$ a vector $f(x)$ in such a way that for any

**Table 2.** Definition and calculation of total (whole-molecule) and local (atom) linear indices of the molecular pseudograph's atom adjacency matrix of the 2-aminobenzaldehyde molecule




2-Amino-benzaldehyde

**Molecular Structure**

**Molecular Pseudograph (G) (Hydrogen Suppressed-pseudograph)**

$\mathbf{X} = [O_1, C_2, C_3, C_4, C_5, C_6, C_7, C_8, N_9]$

Molecular Vector: $\mathbf{X} \in \Re^9$

In the definition of the **X,** as molecular vector, the chemical symbol of the element is used to indicate the corresponding electronegativity value. That is: if I write O it means $\chi$(O), oxygen Mulliken electronegativity or some atomic property, which characterizes each atom in the molecule. Therefore, if I use the canonic basis of $\Re^9$, the coordinates of any vector **X** coincide with the components of that molecular vector

[X] = [3.17, 2.63, 2.63, 2.63, 2.63, 2.63, 2.63, 2.63, 2.33]

[X]: column vector of coordinates of **X** in the Canonical base of $R^9$ (a $nx1$ matrix)

$$f_1(x_i) = \sum_{j=1}^{n} {}^{1}a_{ij} X_j = \mathbf{M^1}[\mathrm{X}] = \begin{bmatrix} 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 2 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} O_1 \\ C_2 \\ C_3 \\ C_4 \\ C_5 \\ C_6 \\ C_7 \\ C_8 \\ N_9 \end{bmatrix} = \begin{bmatrix} 2C_2 \\ 2O_1 + 1C_3 \\ 1C_2 + 1C_3 + 1C_4 + 1C_8 \\ 1C_3 + 1C_4 + 1C_5 \\ 1C_4 + 1C_5 + 1C_6 \\ 1C_5 + 1C_6 + 1C_7 \\ 1C_6 + 1C_7 + 1C_8 \\ 1C_3 + 1C_7 + 1C_8 + 1N_9 \\ 1C_8 \end{bmatrix}$$

Atom linear indices of first order is a *linear map*; $f_1(x_i)$: $\Re^n \rightarrow \Re^n$ *such that,*

$f_1$(O, C, C, C, C, C, C, C, N) = (2C, 2O+1C, 1C +1C+1C+1C, 1C+1C+1C, 1C+1C+1C, 1C+1C+1C, 1C+1C+1C, 1C+1C+1C+1N, 1C) = (5.26, 8.97, 10.52, 7.89, 7.89, 7.89, 7.89, 10.22, 2.63) and whole-molecule linear indices of first order is a *linear functional (linear form)*;

$$f_1(x) = \sum_{i=1}^{n} f_1(x_i) = f_1(\mathrm{O}_1) + f_1(\mathrm{C}_2) + f_1(\mathrm{C}_3) + f_1(\mathrm{C}_4) + f_1(\mathrm{C}_5) + f_1(\mathrm{C}_6) + f_1(\mathrm{C}_7) + f_1(\mathrm{C}_8) + f_1(\mathrm{N}_9) = 69.16$$

***Local (Atom) and total (Whole-Molecule) linear indices of order 0-5*** ($k$ = 0-5)

| Atom ($i$) | $f_0(x_i)$ | $f_1(x_i)$ | $f_2(x_i)$ | $f_3(x_i)$ | $f_4(x_i)$ | $f_5(x_i)$ |
|---|---|---|---|---|---|---|
| $O_1$ | 3.17 | 5.26 | 17.94 | 42.08 | 146.96 | 400.72 |
| $C_2$ | 2.63 | 8.97 | 21.04 | 73.48 | 200.36 | 676.25 |
| $C_3$ | 2.63 | 10.52 | 37.6 | 116.2 | 382.33 | 1193.57 |
| $C_4$ | 2.63 | 7.89 | 26.3 | 87.57 | 277.41 | 894.29 |
| $C_5$ | 2.63 | 7.89 | 23.67 | 73.64 | 234.55 | 739.87 |
| $C_6$ | 2.63 | 7.89 | 23.67 | 73.34 | 227.91 | 721.81 |
| $C_7$ | 2.63 | 7.89 | 26 | 80.93 | 259.35 | 820.73 |
| $C_8$ | 2.63 | 10.22 | 31.26 | 105.08 | 333.47 | 1080.23 |
| $N_9$ | 2.33 | 2.63 | 10.22 | 31.26 | 105.08 | 333.47 |
| **Total** | 23.91 | 69.16 | 217.7 | 683.58 | 2167.42 | 6860.94 |

$$f(\lambda_1 X_1 + \lambda_2 X_2) = \lambda_1 f(X_1) + \lambda_2 f(X_2) \quad (3)$$

scalar $\lambda_1$, $\lambda_2$ and any vector $X_1$, $X_2$ in $\Re^n$. The defining equation 1 for $\boldsymbol{f}_k(x_i)$ may be written as a single matrix equation:

$$\boldsymbol{f}_k(x_i) = [X']^k = \mathbf{M}^k[X] \quad (4)$$

where $[X]$ is a column vector (a $n \times 1$ matrix) of the coordinates of $X$ in the canonical basis of $\Re^n$ and $M^k$ the $k$th power of the matrix $M$ of the molecular pseudograph (map's matrix).

Note that this approach is rather similar to the LCAO-MO (Linear Combinations of Atomic Orbitals- Molecular Orbitals) method. Really, our approach (for $k = 1$) is a quite similar approximation to the Hückel MO method, because in our formalism each MO $\psi_i$ is composed of $n$ valence AOs of atoms in a molecule.

The main idea of the LCAO-MO method is that the electrons in a molecule are accommodated in definite MOs just as in an atom they are accommodated in definite AOs. Normally MOs are made up as LCAO of atoms composing the system, that is, are written in the form,

$$\psi_i = \sum_{j=1}^{n} c_{ij}\varphi_j \quad (5)$$

where $i$ is the number of the MO $\psi$ [in our case, $\boldsymbol{f}_1(x_i)$]; $j$ are the numbers of atomic $\varphi$-orbitals (in our case, $X_j$); $c_{ij}$ (in our case, ${}^1a_{ij}$) are the numerical coefficients defining the contributions of individuals AOs into the given MO. Such a way of constructing a MO is based on the assumption that an atom represented by a definite set of orbitals remains distinctive in the molecule.

Total (whole-molecule) linear indices are *linear functionals* (some mathematicians use the term *linear form*, which means the same as linear functional) on $\Re^n$. That is, the $k$th total linear index is a linear map from $\Re^n$ to the scalar $\Re[\boldsymbol{f}_k(x) : \Re^n \to \Re]$. The mathematical definition of these molecular descriptors is the following:

$$f_k(x) = \sum_{i=1}^{n} f_k(x_i) \quad (6)$$

where $n$ is the number of atoms and $\boldsymbol{f}_k(x_i)$ are the atom's linear indices (linear maps) obtained by Eq. 1. Then, a linear form $\boldsymbol{f}_k(x)$ can be written in matrix form,

$$\boldsymbol{f}_k(x) = [u]^t[X']^k \quad (7)$$

or

$$\boldsymbol{f}_k(x) = [u]^t\mathbf{M}^k[X] \quad (8)$$

for each molecular vector $X \in \Re^n$. $[u]^t$ is a $n$-dimensional unitary row vector. As can be seen, the $k$th total linear index is calculated by summing the local (atom) linear indices of all atoms in the molecule.

In addition to atom linear indices computed for each atom in the molecule, a local-fragment (atom-type) formalism can be developed. The $k$th atom-type linear index of the molecular pseudograph's atom adjacency matrix is calculated by summing the $k$th atom linear indices of all atoms of the same atom type in the molecule. Consequently, if a molecule is partitioned in $Z$ molecular fragments, the total linear indices can be partitioned in $Z$ local linear indices $\boldsymbol{f}_{kL}(x)$, $L = 1, \ldots, Z$. That is to say, the total linear indices of order $k$ can be expressed as the sum of the local linear indices of the $Z$ fragments of the same order:

$$f_k(x) = \sum_{L=1}^{Z} f_{kL}(x) \quad (9)$$

In the atom-type linear indices formalism, each atom in the molecule is classified into an atom-type (fragment), such as heteroatoms (O, N, and S), H-bonding to heteroatoms, halogens atoms, aliphatic carbon chain, aromatic atoms (aromatic rings), and so on. For all data sets, including those with a common molecular scaffold as well as those with very diverse structure, the $k$th fragment (atom-type) linear indices provide much useful information.

## 3. Results and discussion

### 3.1. Development of the classification model

The general data set (see Experimental)[1,3,29,30] was randomly divided into two subsets, training and test set, both of them containing anthelmintic and non-anthelmintic compounds. The active chemicals were selected by considering representatives of most of the different structural patterns and action mechanisms of anthelmintic activity. In this sense, the inactive chemicals were selected in a random way of a great database, with different pharmacological uses. The classification obtained model (to discriminate anthelmintic from non-anthelmintic compounds belonging to the training set) is given below together with the LDA-statistical parameters:

$$\begin{aligned} \text{Class} = {} & 3.30919 + 0.04397\boldsymbol{f}_0(x) - 0.30477\boldsymbol{f}_{1\text{L}}(x_{\text{E}}) \\ & + 0.05313\boldsymbol{f}_{2\text{L}}(x_{\text{E}}) + 0.860526\boldsymbol{f}_{1\text{L}}(x_{\text{E–H}}) \\ & - 1.349\boldsymbol{f}_{3\text{L}}(x_{\text{E–H}}) + 0.28659\boldsymbol{f}_{5\text{L}}(x_{\text{E–H}}) \\ & - 0.0171\boldsymbol{f}_{7\text{L}}(x_{\text{E–H}}) + 3.9229 \times 10^{-6}\boldsymbol{f}_{12\text{L}}(x_{\text{E–H}}) \\ N = 273 \;\; & \lambda = 0.46 \;\; D^2 = 4.68 \\ F(8.264) = 38.529 \;\; & p < 0.0000 \end{aligned} \quad (10)$$

where $N$ is the number of compounds, $\lambda$ is the Wilks' statistic, $D^2$ is the squared Mahalanobis distance and $F$ is the Fisher ratio.

The training set involved 144 active (anthelmintic) compounds and 129 inactive ones. In this set, the obtained model (Eq. 10) has a positive predictive value of 95.14% and 84.50% as negative predictive value, for an

**Table 3.** Results of the classification of compounds in training set

| Compounds | $\Delta P\%$[a] | Compounds | $\Delta P\%$[a] | Compounds | $\Delta P\%$[a] |
|---|---|---|---|---|---|
| *Training active group* | | | | | |
| Tetrachloromethane | 92.86 | Nitroscanate | 43.45 | Lucanthone | 74.02 |
| Perchloroethane | 92.39 | G-572 | 90.37 | Diamphenetide | 93.04 |
| Antimony sodium thioglycollate | 66.33 | Resurantel | 91.03 | Miracil C | 1.12 |
| Tartaric acid | 39.16 | Dichlorophen | 94.76 | Dimantine | 95.75 |
| Chlolorobutane | 93.55 | Phenzidole | 90.75 | Alazanine Triclofenate | 87.32 |
| Disophenol | 43.97 | Cyclobendazole | 65.54 | Closantel | 97.37 |
| Stibomen | 80.23 | Oxantel embonate | 66.55 | Becanthone | 77.08 |
| Lindane | 86.41 | Eliminoxy | 96.83 | Dithiazanine iodide | 89.42 |
| Dimethylpiperazine | 56.85 | Parbendazole | 80.19 | Methylrosanilunium chloride | 59.64 |
| Oltipraz | 43.30 | Amidantel | 64.90 | Bunamide | 97.19 |
| Metyridine | 83.15 | Mirasan | 76.30 | Pyruinium | 92.82 |
| Certuna | 81.14 | SRC-4402 | 77.06 | Agrimophol | 94.54 |
| Butonate | 18.05 | Nocodazole | 52.40 | Desoine | 84.27 |
| Punicine | 87.57 | Tolibenzate | 97.43 | Tymoloverm | 99.46 |
| Antienite | 53.05 | Diphenan | 84.94 | Pretamazium iodide | 94.63 |
| Carbendazim | 57.65 | Haloxon | 66.57 | Chuanliansu | 9.39 |
| Wormin | 92.81 | Cambendazole | 58.71 | Hedaquinium chloride | 97.67 |
| Thiabendazole | 56.49 | Ftalofyne | 89.58 | Dryocrassin | 99.53 |
| Lobendazole | 70.84 | Coralox | 82.54 | Phenithionate | 83.75 |
| Bromothymol | 94.20 | Thiophanate | 68.87 | Abamectin A | 57.92 |
| *p*-Cymine | 97.73 | Evultin | 97.85 | Abamectin B | 33.49 |
| Acidum kainicum | 47.43 | Brotianide | 95.08 | Amocarzine | 93.19 |
| Ascaridole | 90.52 | Afesal | 59.42 | Antimony potassium tartrate | 88.25 |
| Geraniol | 75.19 | Furodazole | 11.96 | Aspidin | 93.24 |
| Eucalyptol | 96.98 | Fenbendazole | 67.93 | Aspidinol | 85.33 |
| R 8231 | 72.61 | Santolactone | 97.24 | 1,4-Bis(trichloromethyl)-bencene | 96.13 |
| Antafenite | 75.21 | Ticarbodine | 95.59 | Bithionol | 92.57 |
| Thiacetaisamide | −50.30 | Butamisole | 82.95 | Carvacrol | 96.37 |
| Tetramisole | 68.47 | Spirazine | −50.92 | Embelim | 97.66 |
| Levamisole | 68.47 | Hydroxypethidine | 70.18 | Epsiprantel | 89.44 |
| Vincofos | 57.51 | Lubisan | 89.35 | Acid filix (b) | 99.17 |
| Bromofenofos | 12.07 | Meclorazepam | 50.06 | Acid filix (c) | 99.06 |
| Feniodium chloride | 96.23 | Flubendazole | 80.59 | Glycarsamide | 58.60 |
| Phenotiazine | 96.97 | RP 12869 | 45.77 | Hectolin | 90.09 |
| Phoxim | −1.80 | Piperamide maleate | 32.09 | Mibbemycin A4 | 88.01 |
| Albendazole | 45.33 | Imcarbofos | −84.47 | Moxidectin | 83.88 |
| Anthiolimine | −32.18 | Etibendazole | 57.83 | Naphthalene | 97.73 |
| Morantel | 70.00 | Agrimonulide | 83.30 | 2-Naphthol | 94.70 |
| Nitazoxanide | −43.70 | Egressin | 91.26 | α-Kosin | 56.30 |
| Bendamizole | 16.23 | Artemether | 76.65 | β-Kosin | 75.01 |
| Tioxidazole | 56.74 | Bephenium hydroxynaphthoate | 83.03 | Mibbemycin A3 | 88.99 |
| Cetovex | 67.91 | Flurantel | 64.95 | Nitroxynil | 36.75 |
| 4-Hexylresorcinol | 95.09 | Amphotaline | 85.22 | Paraherquanude | 84.25 |
| Pexantel | 66.63 | Difesyl | 73.53 | Thymol | 95.11 |
| Butynorate | 96.01 | Praziquantel | 96.07 | Rafoxanide | 96.86 |
| Oxyclozanide | 93.84 | 153 C 51 | 81.39 | α-Santonin | 97.24 |
| Hexachlorophenimonophosphate | 11.05 | Trichlorophen | 95.81 | Tenium Closylate | 63.23 |
| Nirtroclofene | −34.46 | Miracil B | 75.06 | Quinacrine | 64.17 |
| *Training inactive group* | | | | | |
| Anfotericina B | −87.96 | Prontosil | −97.70 | Chlortetracycline | −39.38 |
| Melfalan | −32.97 | Primaquina | −57.51 | Caffeine | −32.74 |
| Busulfan | −73.63 | Dauronobicina | −59.98 | Dipyrrone | −50.24 |
| Estreptozocina | −96.58 | Doxorrubicina | −95.59 | Isoprenalina | −4.11 |
| Clorzotozina | −96.96 | Hydroxyurea | −88.26 | Methaminodiazepoxide | −12.85 |
| Metatrexato | −99.73 | Tolbutamide | −77.54 | Metapyrilene | 23.07 |
| Fluorouracilo | 84.32 | Acetohexamide | −50.75 | Trimethoprin | −81.88 |
| Floxundina | −70.01 | Clorpropamide | −83.22 | Dipyridamole | −100.00 |
| Clorodeoxiundina | −73.72 | IPTD | −92.72 | Theophyline | −13.28 |
| Bromodeoxiundina | −74.63 | Carbutamide | −91.90 | Papaverine | 74.66 |
| Trimidina | −72.57 | Metahexamida | −89.30 | Gentamicina | −99.96 |
| Indoxuridina | −75.37 | Glibenclamida | −96.93 | Ticarcillin | −46.11 |
| Citanabina | −83.75 | Ceftriaxona | −99.97 | Furosemida | −99.01 |
| Azaundina | −97.37 | Gliburida | −94.13 | Clozapina | 83.02 |
| Mercaptopurina | −34.87 | Gliburnurida | −74.13 | Periciazina | 35.12 |

(*continued on next page*)

**Table 3** (*continued*)

| Compounds | $\Delta P\%$[a] | Compounds | $\Delta P\%$[a] | Compounds | $\Delta P\%$[a] |
|---|---|---|---|---|---|
| Nafalina | 6.95 | Vinblastina | −2.68 | Nitrazepan | 31.88 |
| Ampicilina | −76.74 | Azatioprina | −98.36 | Brorizolan | 50.50 |
| Amoxicilina | −85.77 | Ergotamina | 17.60 | Lormetazepan | 34.45 |
| Meticilina | −45.73 | Ciplatino | −92.53 | Hidroxizina | −32.62 |
| Oxacilina | −25.01 | Procarbazina | −95.14 | Zopicolona | −72.27 |
| Dicloxacilina | −44.68 | Vindesina | −38.64 | Piritinol | −90.25 |
| Cloxacilina | −29.68 | Adnamicina | −83.23 | Carbucisteina | −35.83 |
| Floxacilina | −27.25 | Mostasa vacilica | −2.92 | Ganciclovir | −99.93 |
| Cefalofina | 0.08 | Clindamicina | −4.12 | Nifuratel | −71.70 |
| Aurotioglucosa | −56.43 | Aciclovir | −99.81 | Ketoconazol | −22.71 |
| Mephenesin | −94.30 | Alupurinol | −34.07 | Capreomicina | −99.99 |
| Cloranfenicol | −79.32 | Tioguanina | −81.81 | Pentostatina | −99.81 |
| Isoniazida | −98.79 | Mitoxantrona | −99.99 | Vinorelbina | −8.70 |
| Rifampicina | −98.02 | Zalcitabina | −68.73 | Nifenazona | 80.48 |
| Piperacilina | −70.81 | TEM | −71.57 | Ranitidina | −85.51 |
| Nitrofurantoina | −59.26 | Tiotepa | −21.66 | Labetalol | −68.56 |
| Cefazolina | −99.69 | HMN | −86.63 | Atenolol | −61.46 |
| Cefalexina | −78.70 | Pentamethyl melamina | −77.66 | Bisaprolol | −5.86 |
| Cefradina | −55.04 | Carmustina | −52.28 | Carvediol | −76.44 |
| Cefamandol | −99.55 | Dacarbazina | −95.47 | Timolol | −90.10 |
| Dapsona | 42.91 | Pentaquina | 10.93 | Ziduvudina | −99.92 |
| Sulfuxona | −96.75 | Probenecid | 32.49 | Nimodipina | 48.27 |
| Tiosemicarbazona | 12.32 | Melfalano | −32.97 | Dinitrato de Isosurbide | −93.11 |
| Sulfanilamida | −94.57 | Cimetidina | −85.65 | Nitroglicerina | −98.82 |
| Sulfisuxazol | −67.76 | Acebutolol | 38.34 | Tetranitrato | −99.96 |
| Sulfametoxipiridazina | −93.93 | Carbimazol | 40.52 | Hidralazina | −99.37 |
| Sulfapiridina | −49.60 | Primidona | 23.01 | Minoxidilo | −83.27 |
| Sulfaguanidina | −98.90 | Chlorothiazide | −99.98 | Pirazinamida | −26.69 |

[a] $\Delta P\% = [P(\text{Active}) - P(\text{Inactive})] \times 100$ (see Chemometric method).

**Table 4.** Overall measures of accuracy obtained in the training and predictive sets for the obtained model

| | Matthews Corr. coefficient | Accuracy (%) | Sensitivity (%) | Specificity (%) | Predictive value (+) (%) | Predictive value (−) (%) |
|---|---|---|---|---|---|---|
| Training set | 0.80 | 90.11 | 87.26 | 93.97 | 95.14 | 84.50 |
| Test set | 0.77 | 88.55 | 84.21 | 92.22 | 90.14 | 87.37 |

accuracy of 90.11% (correct classification). This model showed a high Matthews' correlation coefficient (MCC) of 0.80; MCC quantified the strength of the linear relation between the molecular descriptors and the classifications.[31] Table 3 depicts the results of the classification for the training set. These result and the two most commonly used *operating characteristics* of 'diagnostic' tests (sensitivity and specificity) are depicted in Table 4.

For a more exhaustive testing of the predictive power of the model found a leave-*n*-out cross-validation procedures was carried out. This model shown an 82.63%, 88.04%, 88.92%, 89.55%, 89.89%, 89.20%, 88.42%, 89.89%, 89.93%, 89.51%, 89.16%, 88.42%, 88.38%, 88.77%, 88.38%, 88.38%, 88.77%, 89.16%, and 88.38% of global good classification when *n* varied from 2 to 20 in the leave-*n*-out cross-validation procedures (see Fig. 1).

An important point of view to accept or reject a model as Eq. 10, are the statistics for the external prediction set.[32] The discriminant model correctly classifies 90.14% (positive predictive value) of active

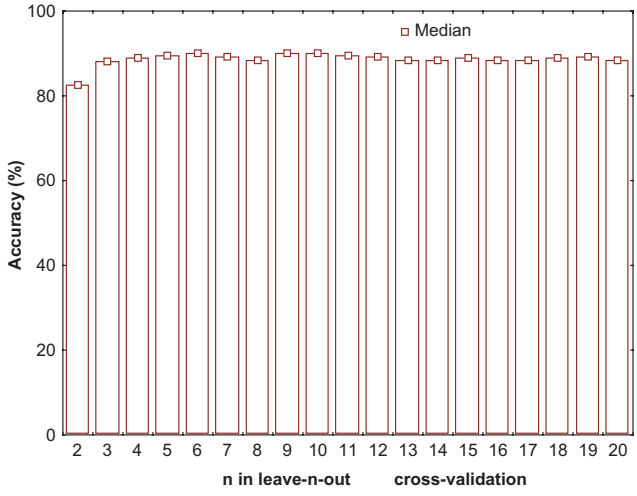



**Figure 1.** Behavior of the overall accuracy (Eq. 10) in jackknife (leave-*n*-out cross-validation) experiment.

**Table 5.** Results of the classification of compounds in an external prediction set

| Compounds | $\Delta P\%$[a] | Compounds | $\Delta P\%$[a] | Compounds | $\Delta P\%$[a] |
|---|---|---|---|---|---|
| *Test active group* | | | | | |
| Tetrachloroethylene | 92.57 | Oxfendazole | 8.95 | Acid filix (a) | 98.92 |
| Dichlorvos | 17.85 | Dribendazole | 62.97 | Gentian violet | 68.20 |
| Fospirate | 13.49 | Domoic acid | 35.88 | Clorsulon | −100.00 |
| Piperazine | −50.01 | Mebendazole | 75.24 | 8-Hydroxyquinoleine | 82.16 |
| Bitoscanate | 70.43 | Benacyl | 66.24 | Dichlorophenersine | 56.79 |
| Methylarecaidine | 70.62 | Santoperonim | 99.50 | Hycanthone | −34.68 |
| Sodium antimony dimethylcysteine tartrate | 96.31 | Tetrachlorodifluoroethane | 94.68 | Diaamphenethide deacetylated (metabolite) | 49.27 |
| Triclabendazole sulfoxide | 89.24 | Pararosaniline embonate | 91.87 | Mandelic acid isoamyl ester | 88.70 |
| Iodothymol | 94.21 | Bemosat | 81.57 | Naftalofos | 32.51 |
| Diethylcarbamazine | 6.91 | Salantel | 96.13 | Arsamilate | 94.20 |
| Pyrantel | 66.93 | Febantel | 62.85 | Stibophen | −100.00 |
| Niridazole | −85.28 | Miracil A | 76.94 | Kainic acid | 47.43 |
| Oxibendazole | 45.46 | Desaspidin | 97.61 | Hexachlorophene | 94.92 |
| Carbantel | 98.29 | Dicroden | 43.55 | Dibromsalam | 93.80 |
| Crufomate | 6.60 | Bidimazium iodide | 88.79 | F 30066 | 0.63 |
| Niclosamide | 52.04 | RO 2-9009 | 68.97 | Dibutyltin dilarate | 99.92 |
| Amoscanate | 62.83 | Oxamniquine | −87.10 | Ivermectin | 88.23 |
| Netobimin | −99.68 | Bisbendazole | 42.22 | Doramectin | 54.15 |
| Ontianil potassium | 95.74 | Teroxalene | 93.34 | Milbemycin D | 94.66 |
| Atractyl | 85.18 | Stilbazium iodide | 92.96 | Dymanthine | 95.75 |
| Triclabendazole | 79.80 | Artesunate | 73.92 | Tribromsalam | 93.66 |
| Styrylpyridinium chloride | 93.77 | Alantolactone | 97.80 | Bromoxanide | 79.52 |
| Coomafos | 81.64 | Arecoline | 70.62 | Tetrachloroethane | 87.36 |
| Clioxanide | 93.98 | Urea stibamine | 59.71 | | |
| *Test inactive group* | | | | | |
| Trifluoromethyl deoxiundina | −51.68 | Sucrida | −96.60 | Carbenicillin | −21.06 |
| Azanibina | −96.38 | Mitomicina | −82.26 | Diflunisal | 92.79 |
| Carbenicilina | −21.06 | Ciclofosfamida | −12.26 | Loprazolan | −78.73 |
| Cefapirina | −69.00 | $PGE_1$ | 65.42 | Chlorpromazine | 76.94 |
| Cefoxilina | −75.26 | Epinubicina | −99.30 | Pentamidina | −66.23 |
| Phenobarbital | 81.95 | TEPA | −22.90 | Metronidazol | −88.57 |
| Halazepan | 96.85 | Etoposido | −20.56 | Propatilnitrato | −99.00 |
| Sulfametazina | −83.79 | Difenilhidramina | 88.28 | Ac.mefenamico | 96.56 |
| Sulfatiazol | −80.25 | Novobiocin | −24.90 | Clofibrato | 94.85 |
| Bufeina | 85.52 | Amrinona | −22.96 | Supazine | −99.75 |
| Mannitol | −88.23 | Lamotrigine | −90.68 | Mepartircin B | −99.77 |
| Gentiopicrim | −68.62 | Homofasalazine | −85.96 | Peralopride | −15.67 |
| Sulfametoxazol | −82.74 | Phenytoin | 88.35 | Teniposiclo | −60.10 |
| Carteolol | 66.84 | Benazeprilo | 87.74 | Disopiramina | −99.93 |
| Sulfacetamida | −66.43 | Tetraciclina | −47.28 | Procainamida | −1.21 |
| Sulfadiazina | −82.02 | Acetazolamida | −99.64 | Mononitrato de liusurbida | −75.22 |
| Gossypol | −88.69 | Cisapride | −3.63 | Mepartricin A | −99.89 |
| Tolazamida | −74.67 | Disulfiran | −16.82 | Cycloquin | −37.79 |
| Sucrose | −99.88 | Acido oxidronico | −95.81 | *N*-Methylglucamine | −95.58 |
| Sumatriptan | −96.54 | Acarbose | −100.00 | Narceina | −49.48 |
| Taurocholic acid | −98.10 | Gramsetron | −30.73 | Folcodina | −49.25 |
| Gly-Pro | −83.71 | Glucin | −93.24 | Etonitazene | −14.85 |
| D-Glucose | −66.19 | Saccharin | −69.88 | Gobad | −67.52 |
| Raffinose | −99.99 | Urapidil | −57.07 | Tironamine | −31.23 |
| Sidenafil | −98.24 | Cycloserine | −3.55 | Fluopsin C | −39.53 |
| Lactulose | −99.84 | Loracarbef | −84.79 | Taurultam | −98.90 |
| Bosentan | −98.53 | Flurnithine | −19.20 | Noxytiolin | −78.44 |
| Glycine | −93.20 | AAFC | −56.88 | Acivicin | −10.83 |
| Phe-Pro | −40.86 | Xamoterol | −90.52 | Thiram | −59.19 |
| Gabapentin | −48.53 | Gliclazide | −85.25 | Metenamine | −80.66 |
| BVARAU | −92.43 | Cromolym sodium | −38.17 | Diamide | −48.04 |
| Lucifer yellow | −100.00 | Lissoclinotoxin A | −99.46 | | |

[a] $\Delta P\% = [P(\text{Active}) - P(\text{Inactive})] \times 100$ (see Chemometric method).

compounds and 87.37% (negative predictive value) of inactive compounds in the test set, this model showed a Matthews' correlation coefficient (MCC) of 0.77 (88.55% of correct classification). In Table 5, we give the classification of compounds in this set.

### 3.2. Development of the discriminant functions to predict the mode of action of anthelmintic drugs

In order to develop a QSAR model that permits to predict the main modes of action of the anthelmintic drugs selected as active with the use of the discriminant model (Eq. 10), we used a series of five more representative and studied action modes of the anthelmintic compounds. The targets selected to develop this in silico study contain three pharmacological targets 'coupled' to ions channels and other two unrelated action places (at least directly) with ions channels (see General data set). The linear classification functions to describe the main modes of action of the anthelmintic drugs are given below together with their statistical parameters:

$$\begin{aligned}&\text{Nicotine AcetylChol. Agonist (NAA)}\\&= -5.3963 + 0.20257\boldsymbol{f}_0(x)\\&\quad - 0.3519\boldsymbol{f}_{0\mathrm{L}}(x_\mathrm{E})0.01745\boldsymbol{f}_{2\mathrm{L}}(x_\mathrm{E})\end{aligned} \tag{11}$$

$$\begin{aligned}&\text{AcetylCholinest. Inh. (AChI)}\\&= -17.3075 + 0.057155\boldsymbol{f}_0(x)\\&\quad - 0.44064\boldsymbol{f}_{0\mathrm{L}}(x_\mathrm{E}) + 0.160097\boldsymbol{f}_{2\mathrm{L}}(x_\mathrm{E})\end{aligned} \tag{12}$$

$$\begin{aligned}&\text{GluCl Receptor (GluClR)}\\&= -33.5675 + 0.488608\boldsymbol{f}_0(x)\\&\quad + 0.224602\boldsymbol{f}_{0\mathrm{L}}(x_\mathrm{E}) - 0.07562\boldsymbol{f}_{2\mathrm{L}}(x_\mathrm{E})\end{aligned} \tag{13}$$

$$\begin{aligned}\beta\text{-Tubulin (t)} &= -8.03235 + 0.119206\boldsymbol{f}_0(x)\\&\quad - 0.25502\boldsymbol{f}_{0\mathrm{L}}(x_\mathrm{E})\\&\quad + 0.072003\boldsymbol{f}_{2\mathrm{L}}(x_\mathrm{E})\end{aligned} \tag{14}$$

$$\begin{aligned}\text{H-Ionophores (H)} &= -9.5741 + 0.083876\boldsymbol{f}_0(x)\\&\quad + 0.77098\boldsymbol{f}_{0\mathrm{L}}(x_\mathrm{E})\\&\quad - 0.06161\boldsymbol{f}_{2\mathrm{L}}(x_\mathrm{E})\\N = 73\ \lambda = 0.033\ F(12,174) &= 38.193\\p < 0.0000\end{aligned} \tag{15}$$

We use a data set of 93 compounds, for which a specific mode of action has been reported. It was randomly split in two subsets, one of them (training set) with 73 compounds and the other (test set) with 20 compounds. The obtained model achieved an accuracy of 93.50% in the training set and 80.00% in the test set. Table 6 shows the squared Mahalanobis distance ($D^2$) between the groups and the classification matrix for the training set and we show the graphical results of canonical analysis in Figure 2. In addition, Table 7 depicts the results for the both series (training and test sets) when the discriminant function was used.

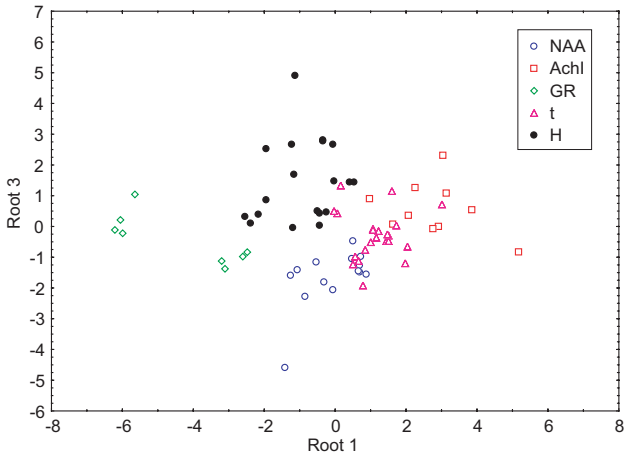



**Figure 2.** Graphical representation of the results of the canonical analysis.

### 3.3. Virtual screening of anthelmintic compounds

The massive cost of developing new drugs, coupled with candidate attrition rates during the discovery and development process highlights the need for a 'sea change' in the drug discovery paradigm. Predictive in silico models could be used for the desired property identification, accelerating the selection process of lead compounds and predicting their modes of action.[33] One of the most important features of any QSAR model is its ability to predict the desired property for new compounds from databases of chemicals.[17] Computational 'in silico' screening of large databases considering the use of such models has emerged as an interesting alternative to HTS and an important drug-discovery tool.[34,35] With the aim of prove the possibilities of the TOMOCOMD-CARDD approach to detect new leads, we performed a simulated virtual screening of 10 compounds reported in the medical literature as promissory anthelmintic [36–38] whose structures are given in Table 8.

**Table 6.** Squared Mahalanobis distance ($D^2$) among groups (modes of action) and percentage of good classification (%) for the training set

| | (NAA) | (AchI) | (GR) | (t) | (H) | | % | (NAA) | (AchI) | (GR) | (t) | (H) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| (NAA) | 0 | 23.77 | 33.51 | 5.45 | 11.10 | (NAA) | 100 | 14 | 0 | 0 | 0 | 0 |
| (AchI) | | 0 | 56.65 | 6.74 | 24.56 | (AchI) | 70 | 0 | 7 | 0 | 3 | 0 |
| (GR) | | | 0 | 37.35 | 27.84 | (GR) | 100 | 0 | 0 | 8 | 0 | 0 |
| (t) | | | | 0 | 9.33 | (t) | 90.90 | 1 | 1 | 0 | 20 | 0 |
| (H) | | | | | 0 | (H) | 100 | 0 | 0 | 0 | 0 | 19 |
| | | | | | | Total | 93.15 | 15 | 8 | 8 | 23 | 19 |

**Table 7.** Results of the classification of anthelmintic compounds for the training and test series

| Compound[a] | (NAA)[b] | (AChI)[b] | (GR)[b] | (t)[b] | (H)[b] | Compound[a] | (NAA)[b] | (AChI)[b] | (GR)[b] | (t)[b] | (H)[b] |
|---|---|---|---|---|---|---|---|---|---|---|---|
| *Nicotinic acetylcholine agonist* (*NAA*) | | | | | | | | | | | |
| Metyridine | 90.26 | 0.00 | 0.00 | 7.86 | 1.88 | Butamisole | 51.28 | 0.17 | 0.00 | 47.97 | 0.58 |
| R 8231 | 67.16 | 0.03 | 0.00 | 31.23 | 1.57 | Bephenium hydroxynaphthoate | 98.82 | 0.00 | 0.00 | 1.02 | 0.16 |
| Levamisole* | 83.80 | 0.01 | 0.00 | 15.76 | 0.43 | Bemosat | 93.06 | 0.00 | 0.00 | 4.32 | 2.62 |
| Antafenite | 79.73 | 0.02 | 0.00 | 19.98 | 0.27 | Difesyl | 94.78 | 0.00 | 0.01 | 3.38 | 1.83 |
| Tetramisole | 83.80 | 0.01 | 0.00 | 15.76 | 0.43 | Morantel* | 87.21 | 0.01 | 0.00 | 12.43 | 0.36 |
| Pyrantel | 84.52 | 0.01 | 0.00 | 15.00 | 0.47 | Stilbazium iodide | 99.18 | 0.00 | 0.28 | 0.55 | 0.00 |
| Oxantel embonate | 93.39 | 0.00 | 0.00 | 3.40 | 3.21 | Arecoline | 73.41 | 0.00 | 0.00 | 17.31 | 9.28 |
| Styrylpyridinium chloride | 96.57 | 0.00 | 0.00 | 3.28 | 0.15 | Tenium Closylate | 95.75 | 0.00 | 0.00 | 3.84 | 0.40 |
| Antienite* | 51.17 | 0.16 | 0.00 | 47.84 | 0.83 | | | | | | |
| *Acetylcholinesterase inhibitors (AChI)* | | | | | | | | | | | |
| Trichlorfon* | 0.05 | 0.06 | 0.00 | 2.99 | 96.90 | Coralox | 0.01 | 90.46 | 0.00 | 9.53 | 0.00 |
| Haloxon* | 0.01 | 85.16 | 0.00 | 14.70 | 0.13 | Imcarbofos | 0.00 | 99.87 | 0.00 | 0.13 | 0.00 |
| Fospirate | 0.25 | 34.85 | 0.00 | 63.13 | 1.77 | Butonate* | 0.10 | 78.07 | 0.00 | 19.63 | 2.20 |
| Famophos | 0.00 | 98.31 | 0.00 | 1.69 | 0.00 | Uredofos | 0.00 | 100.00 | 0.00 | 0.00 | 0.00 |
| Dichlorvos* | 0.67 | 50.02 | 0.00 | 46.80 | 2.51 | Diuredosan | 0.00 | 99.81 | 0.00 | 0.19 | 0.00 |
| Vincofos | 3.92 | 1.47 | 0.00 | 67.94 | 26.67 | Zilantel | 0.00 | 99.42 | 0.00 | 0.58 | 0.00 |
| Coomafos | 0.03 | 83.67 | 0.00 | 16.30 | 0.00 | Crufomate* | 14.77 | 0.86 | 0.00 | 78.56 | 5.81 |
| Metrifonate* | 0.05 | 0.06 | 0.00 | 2.99 | 96.90 | Naftalofos | 0.67 | 31.71 | 0.00 | 67.39 | 0.23 |
| *GluCl receptor* (*GR*) | | | | | | | | | | | |
| Abamectin A | 0.00 | 0.00 | 100.00 | 0.00 | 0.00 | Moxidectin | 0.02 | 0.00 | 99.97 | 0.00 | 0.01 |
| Abamectin B | 0.00 | 0.00 | 100.00 | 0.00 | 0.00 | Ivermectin | 0.00 | 0.00 | 100.00 | 0.00 | 0.00 |
| Mibbemycin A3 | 1.87 | 0.00 | 96.81 | 0.36 | 0.95 | Doramectin | 0.00 | 0.00 | 100.00 | 0.00 | 0.00 |
| Mibbemycin A4 | 3.80 | 0.00 | 92.64 | 0.92 | 2.64 | Milbemycin D | 0.85 | 0.00 | 98.93 | 0.04 | 0.18 |
| β-*Tubulin binding* (*t*) | | | | | | | | | | | |
| Carbendazm | 18.05 | 0.87 | 0.00 | 79.00 | 2.08 | Furodazole | 46.01 | 0.21 | 0.00 | 52.94 | 0.84 |
| Thiabendazole* | 46.70 | 0.23 | 0.00 | 52.71 | 0.37 | Oxfendazole | 3.69 | 9.91 | 0.00 | 86.15 | 0.26 |
| Lobendazole | 19.33 | 0.92 | 0.00 | 78.42 | 1.32 | Dribendazole | 23.16 | 0.83 | 0.00 | 74.89 | 1.11 |
| Bendamizole | 14.98 | 2.93 | 0.00 | 82.03 | 0.07 | Flubendazole | 27.97 | 0.68 | 0.00 | 70.46 | 0.88 |
| Oxibendazole | 13.69 | 1.10 | 0.00 | 81.27 | 3.94 | Fenbendazole* | 20.73 | 1.08 | 0.00 | 77.38 | 0.81 |
| Mebendazole* | 21.82 | 0.94 | 0.00 | 76.15 | 1.09 | Etibendazole | 6.61 | 0.71 | 0.00 | 58.66 | 34.02 |
| Cyclobendazole | 12.50 | 1.70 | 0.00 | 84.21 | 1.59 | Febantel | 0.55 | 7.51 | 0.00 | 56.23 | 35.71 |
| Nocodazole | 5.94 | 4.07 | 0.00 | 88.49 | 1.51 | Benomyl | 3.85 | 8.15 | 0.00 | 87.52 | 0.49 |
| Tioxidazole | 13.41 | 1.12 | 0.00 | 81.42 | 4.05 | Benacyl | 6.68 | 4.77 | 0.00 | 87.88 | 0.67 |
| Parbendazole | 67.96 | 0.04 | 0.00 | 31.12 | 0.88 | Bisbendazole | 2.41 | 32.20 | 0.00 | 65.39 | 0.00 |
| Triclabendazole* | 11.65 | 0.06 | 0.00 | 27.00 | 61.30 | Luxabendazole | 0.00 | 97.80 | 0.00 | 2.20 | 0.00 |
| Cambendazole | 1.93 | 21.42 | 0.00 | 76.60 | 0.05 | Triclabendazole sulfoxide | 12.46 | 0.18 | 0.00 | 44.43 | 42.93 |
| Thiophanate | 0.19 | 40.31 | 0.00 | 57.47 | 2.03 | Albendazole sulfoxide | 2.56 | 10.22 | 0.00 | 86.53 | 0.69 |
| Albendazole* | 15.07 | 1.17 | 0.00 | 81.48 | 2.28 | | | | | | |
| *Proton ionophores* (*H*) | | | | | | | | | | | |
| Disophenol | 1.96 | 0.01 | 0.00 | 6.51 | 91.52 | Ticarbodine | 14.04 | 0.01 | 0.00 | 16.04 | 69.91 |
| Bromofenofos | 0.02 | 0.01 | 0.00 | 1.10 | 98.86 | Flurantel | 0.00 | 0.00 | 0.00 | 0.02 | 99.98 |
| Niclofolan* | 0.01 | 0.01 | 0.00 | 0.49 | 99.49 | Clioxanide* | 12.33 | 0.11 | 0.00 | 36.17 | 51.39 |
| OK sinid | 0.01 | 0.01 | 0.00 | 0.54 | 99.44 | Salantel | 1.42 | 0.00 | 0.06 | 0.57 | 97.95 |
| Oxyclozanide | 0.01 | 0.00 | 0.00 | 0.04 | 99.94 | Trichlorophen | 4.91 | 0.00 | 0.18 | 0.20 | 94.70 |

(*continued on next page*)

**Table 7** (*continued*)

| Compound[a] | (NAA)[b] | (AChI)[b] | (GR)[b] | (t)[b] | (H)[b] | Compound[a] | (NAA)[b] | (AChI)[b] | (GR)[b] | (t)[b] | (H)[b] |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Niclosamide | 0.99 | 0.05 | 0.00 | 9.01 | 89.95 | Closantel | 4.51 | 0.00 | 0.55 | 0.96 | 93.99 |
| Hexachlorophenimonophosphate* | 0.00 | 0.00 | 0.00 | 0.03 | 99.97 | Dibromsalam* | 43.62 | 0.00 | 0.00 | 14.20 | 42.18 |
| Nirtroclofene | 0.01 | 0.01 | 0.00 | 0.54 | 99.44 | Desaspidin | 8.46 | 0.00 | 4.05 | 1.29 | 86.20 |
| Resurantel | 36.39 | 0.00 | 0.00 | 12.18 | 51.42 | Bithionol | 0.23 | 0.00 | 0.00 | 0.17 | 99.60 |
| Dichlorophen | 32.93 | 0.00 | 0.00 | 1.77 | 65.30 | Bromoxanide* | 0.23 | 0.00 | 0.01 | 0.24 | 99.52 |
| Rafoxanide* | 21.19 | 0.01 | 0.02 | 11.72 | 67.06 | Hexachlorophene | 0.01 | 0.00 | 0.00 | 0.01 | 99.98 |
| Brotianide | 12.28 | 0.04 | 0.00 | 24.03 | 63.65 | Tribromsalam | 12.10 | 0.00 | 0.00 | 7.31 | 80.58 |
| Afesal | 0.88 | 1.66 | 0.00 | 39.36 | 58.09 | | | | | | |

[a] Compounds marked with (*) were used in the external test set.
[b] Percentage of probability with which the drug is predicted to be anthelmintic having nicotinic acetylcholine agonist (NAA), acetylcholinesterase inhibitors (AChI), GluCl receptor (GR), β-tubulin inhibitors (t), or proton ionophores (H) as mechanism of action, respectively, using Eqs. (11)–(15).

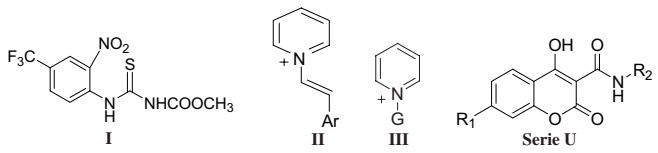
The first compound (**I**) it is a new pro-drug of the benzimidazole type, which has shown activity 'in vivo' against *Echinococcus*.[36] This compound was classified as anthelmintic by the obtained model. The compounds **II**–**III** were taken from the medical literature of the last decade.[37] Most of these compounds have been recognized as anthelmintic and classified as cholinergic agonist, which is coherent if we take into consideration the structural features for the cholinergic agonist activity and the modes of action of drugs like pirantel, morantel, and oxantel which have similar structure. The last four compounds (**U-series**) belong to the family of the β-ketoamides.[38] These compounds have shown a potent activity against nematodes (~1 μM against *H. contortus*); they have been predicted as anthelmintic and classified as oxidative phosphorylase uncoplers (H-ionophores).

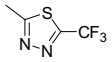
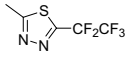
After that, we use some compounds obtained from the literature [39–41] the structures of this compounds are shown in Figure 3. The compound (**IV**), named Dioxapyrrolomycin[39] was classified as active by Eq. 11 and was predicted as oxidative phosphorylase uncoplers (H-ionophores). It could be related with the protonic H-atom in the pyrrol ring. The fifth set (**V**; a new family of cyclic depsipeptides)[40] was correctly selected as anthelmintic by the discriminant function. Finally, a *subsystem* of 5-substituted pyrimidine derivatives (**VI**) was evaluated. These compounds have been reported as trematocide (mainly fasciolicide) and nematocide;[41] and have been predicted as anthelmintic by the obtained model (Eq. 10). Most of them were classified as inhibitor of β-tubulin polymerization. However, this result is preliminary and must be corroborated experimentally. Table 9 shown results when the discriminant function was used.

No compound with this kind of structure was included in the training data set and in this sense; its evaluation is equivalent to the discovery of new lead compounds. Evidently, the quality of these predictions is not accidental, because the calculations of these descriptors not only give explanation of the molecule in a global way, but also certain subfragments are considered too, which are structural features responsible for the activity. These results confirm that the TOMOCOMD-CARDD descriptors used in the work are able to a priori selection/identification and to predict the mode of action of new lead anthelmintics, starting from a set consisting of structurally different compounds.

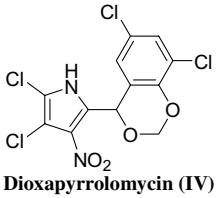
On the other hand, the pharmaceutical industry as a whole embraced this novel paradigm for drug discovery, based on the HTS, as solutions to the time and cost-cutting problem, even though no guarantees for success were provided.[42] While the impact of HTS has yet to be established, the drug discovery rate did not improve in the past decade. In fact, the number of new chemical entities (NCEs) has remained rather constant—an average of 37 NCEs per year—between 1990 and 2000, even though the number of compounds is being initially screened has increased by several orders of magnitude.[43,44]

**Table 8.** Results of the evaluation of compounds reported as anthelmintic

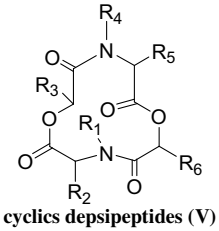



| No or code | Structure Ar | MED[a] (mg/kg) | $\Delta P\%$[b] | (NAA)[c] | (AchI)[c] | (GR)[c] | (t)[c] | (H)[c] |
|---|---|---|---|---|---|---|---|---|
| **I** | — | | 76.79 | 0.01 | 0.01 | 0.00 | 0.51 | 99.48 |
| **II-1** | $C_6H_5$ | 125 | 88.61 | 95.04 | 0.00 | 0.00 | 4.88 | 0.08 |
| **II-2** | *o*-$CH_3C_6H_4$ | 15 | 89.66 | 95.77 | 0.00 | 0.00 | 4.17 | 0.06 |
| **II-3** | *o*-$ClC_6H_4$ | 31 | 87.66 | 92.52 | 0.00 | 0.00 | 7.01 | 0.47 |
| **III-1** | *N*-Allyl | 500 | 81.44 | 90.46 | 0.00 | 0.00 | 9.22 | 0.32 |
| **III-2** | *N*-Methyl | 500 | 75.46 | 88.48 | 0.00 | 0.00 | 10.93 | 0.59 |
| | Structure | Act. Obs. | | | | | | |
| | R1 R2 | | | | | | | |
| **U-86996** | H *p*-$BrC_6H_5$ | | 90.72 | 26.36 | 0.00 | 0.00 | 13.79 | 59.85 |
| **U-87407** | Cl *p*-$BrC_6H_5$ | ~1 μM | 89.70 | 6.60 | 0.00 | 0.00 | 5.10 | 88.29 |
| **U-88509** | Cl  | | 66.67 | 0.00 | 0.00 | 0.00 | 0.14 | 99.86 |
| **U-89605** | Cl  | | 83.50 | 0.00 | 0.00 | 0.00 | 0.00 | 100.00 |

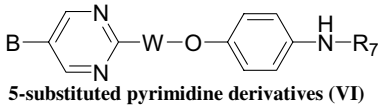
[a] Anthelmintic activity measured experimentally.
[b] Anthelmintic activity predicted by Eq. 10; $\Delta P\% = [P(\text{Active}) - P(\text{Inactive})] \times 100$ (see Chemometric method).
[c] Percentage of probability with which the drug is predicted to be anthelmintic having NAA, AChI, GR, t, H as mechanism of action, respectively, using Eqs. (11)–(15).




**Dioxapyrrolomycin (IV)**
**(LL-F42248)**




**cyclics depsipeptides (V)**




**5-substituted pyrimidine derivatives (VI)**

**Figure 3.** Basic molecular backbone of compounds reported in anthelmintic patent.

At present, dataset of drugs with some pharmacological uses can be assayed in the virtual screening process. This approach is very interesting, because compounds selected with this in silico procedure have well-established methods of synthesis; although in many cases their toxicological, pharmacodynamical and pharmaceutical behaviors are well known. For this reason, we have selected this method of search for the novel anthelmintic compounds.

We had performed an exhaustive search in the Merck Index,[30] Negwer handbook,[29] and Goodman and Gilman[45] looking for compounds to be evaluated in the models. A reduced group was identified by the discriminant function as a possible anthelmintic. Well-known drugs, with other pharmacological properties, were selected as possible anthelmintic by the obtained model. These compounds identified as active, but they have not been reported in the literature as anthelmintic, are now in assay in order to prove their anthelmintic activity experimentally.

Interestingly, one of the compounds selected as anthelmintic was colchicine ($\Delta P\% = 70.57$), which unites to the β-tubulin blocking the polymerization and the formation of the microtubules.[46] Although there are no reports for the use of colchicine as anthelmintic, it has been demonstrated experimentally that anthelmintic benzimidazole competed with the binding site for colchicine in β-tubulin.[3]

### 3.4. Experimental versus theoretical result

Current research in the Chemical Bioactive Center has been focused on the microcidal potential of 2-furylethylenes. For example, **G-1** [*Z*-1-bromo-1-nitro-2-(5-bromofur-2-yl)ethylene] has shown a pronounced double effect (antifungal/antibacterial).[47] It was therefore of major interest for us to develop a general model for predicting the anthelmintic activity for a given molecular structure, with the major emphasis on the case of 2-furylethylenes.

**Table 9.** Results of the classification of compounds included in anthelmintic patents

| Compound (no or code)[a] | | | | | $\Delta P$%[b] | (NAA)[c] | (AchI)[c] | (GR)[c] | (t)[c] | (H)[c] |
|---|---|---|---|---|---|---|---|---|---|---|
| (**IV**) | | | | | 20.83 | 0.05 | 0.00 | 0.00 | 0.57 | 99.38 |
| (**V**) Substituents | | | | | | | | | | |
| $R_1$ and $R_4$ | $R_2$, $R_3$, $R_5$, and $R_6$ | | | | | | | | | |
| Ethyl | Ethyl | | | | 85.81 | 2.00 | 13.26 | 0.00 | 84.09 | 0.65 |
| Propyl | propyl | | | | 92.94 | 26.28 | 0.57 | 1.23 | 71.56 | 0.36 |
| *p*-Br-phenyl | $CH_2Cl$ | | | | 52.58 | 0.02 | 0.04 | 0.26 | 1.38 | 98.30 |
| phenyl | $(CH_2)_2Cl$ | | | | 75.11 | 0.53 | 0.15 | 19.21 | 12.04 | 68.07 |
| Cyclo-propyl | Ethyl | | | | 90.18 | 2.06 | 15.49 | 0.00 | 82.15 | 0.30 |
| Cyclo-pentyl | Ethyl | | | | 93.37 | 4.21 | 8.80 | 0.04 | 86.73 | 0.22 |
| Ethyl | $CH_2Cl$ | | | | 27.64 | 0.02 | 0.06 | 0.00 | 1.95 | 97.97 |
| (**VI**) Substituents | | | Doses (mg/kg) | Post mortem no adults *Fascioles* | | | | | | |
| B | W | $R^7$ | | | | | | | | |
| Cl | — | H | 35 | 0 | 42.04 | 43.13 | 0.04 | 0.00 | 49.64 | 7.19 |
| Cl | — | $CH_3CO$– | 35 | 0 | 63.57 | 15.87 | 0.62 | 0.00 | 80.77 | 2.73 |
| Br | — | H | — | — | 41.29 | 44.44 | 0.05 | 0.00 | 50.09 | 5.42 |
| Cl | –$OCH_2CH_2$– | $CH_3CO$– | 60 | 2 | 42.41 | 13.37 | 0.46 | 0.00 | 76.18 | 9.98 |
| Cl | –$(OCH_2CH_2)_2$– | $CH_3CO$– | 75 | 5 | 13.91 | 9.40 | 0.29 | 0.00 | 59.92 | 30.40 |
| Cl | –$O(CH_2)_3$– | $CH_3CO$– | 35 | 0 | 46.72 | 15.54 | 0.39 | 0.00 | 75.07 | 9.00 |
| Br | –$O(CH_2)_3$– | $CH_3CO$– | 100 | 0 | 46.02 | 16.18 | 0.42 | 0.00 | 76.55 | 6.85 |
| $CF_3$ | –$O(CH_2)_3$– | $CH_3CO$– | 50 | 0 | 87.99 | 0.16 | 0.02 | 0.00 | 3.34 | 96.49 |
| $CH_3$ | –$O(CH_2)_3$– | $CH_3CO$– | 100 | 0 | 54.09 | 25.90 | 0.30 | 0.00 | 71.82 | 1.98 |
| I | –$O(CH_2)_3$– | $CH_3CO$– | 100 | 0 | 45.59 | 16.66 | 0.44 | 0.00 | 77.34 | 5.56 |
| Cl | –$O(CH_2)_4$– | $CH_3CO$– | 50 | 1 | 50.80 | 17.97 | 0.33 | 0.00 | 73.62 | 8.08 |
| Cl | — | =C=S | 50 | 0 | 8.16 | 14.23 | 0.98 | 0.00 | 83.64 | 1.15 |

[a] The chemical substituents of the compounds represented with numbers or codes, are shown in Table 9 and the molecular structure or the backbone are shown in Figure 3.

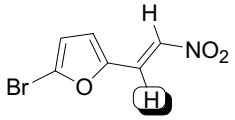
[b] Anthelmintic activity predicted by Eq. 10; $\Delta P = [P(\text{Active}) - P(\text{Inactive})] \times 100$ (see Chemometric method).

[c] Percentage of probability with which the drug is predicted to be anthelmintic having NAA, AChI, GR, t, H as mechanism of action, respectively, using Eqs. 11–15.

In this sense, good agreement between the experimental anthelmintic (effectivity around of 67% at 45 mg/kg against *Fasciola hepatica*) activity test and the predicted activities ($\Delta P = 51.70\%$) for compound **1** [1-nitro-2-(5-bromo-fur-2-yl)ethylene] (see Table 10) was found. As Tables 10 and 11 depicts, compound **1** was identified as anthelmintic (potentially active) $\Delta P = 51.70\%$ by the obtained model (Eq. 10) and had an effectivity around of 67% at 45 mg/kg against *Fasciola hepatica*. So we can say that chemical **1** could be a new anthelmintic lead compound. The result can be considered as a very promissory starting point for the future design and refinement of novel chemicals with higher anthelmintic activity. In this sense, we are looking forward at improving this result in the future by finding more potent candidates. However, our current results are significant because they demonstrate the straightforward way in which molecular linear indices can identify new drug candidates, particularly anthelmintic.

Additionally, chemical **1** was predicted as oxidative phosphorylase uncoplers ($P = 0.54\%$). This is a logic result, because each anthelmintic molecule with this mode of action (proton ionophores) possesses a detachable

**Table 10.** Results in the treatment on *Fasciola hepatica* two week old with compound **1** in Balb/c mice




| P | Dose (mg/kg) | Number *Fasciola* mean (±SD) | IE (%) | Weight gain (g) mean (±SD) | *M* (%) | Affected liver (%) | Hepatic index mean (±SD) | Splein relative weight (g) mean (±SD) | Effectivity *E* (%) |
|---|---|---|---|---|---|---|---|---|---|
| **1** | 16 | 0.86 (0.58) | 86 | −0.05 (0.13) | 0 | 86 | 6.49 (0.26) | 1.1 (0.02) | 0 |
| | 45 | 0.33 (0.37) | 33 | −0.05 (0.01) | 57 | 33 | 6.5 (0.1) | 1.15 (0.01) | 67 |
| Infected | | 0.5 (1.7) | 50 | −1.25 (0.18) | 14 | 50 | 7.24 (0.05) | 1.05 (0.02) | — |
| | | 1 (0.54) | 33 | 0.78 (0.08) | 63 | 33 | 5.8 (0.1) | 0.99 (0.03) | — |
| Control no infected | | | | 1.48 (0.07) | — | — | 5.8 (0.17) | 0.46 (0.02) | — |
| | | | | 0.89 (0.1) | — | — | 4.5 (0.17) | 0.37(0.02) | — |

The location of the dissociating proton (H) is shown in the shaded box. *P* = Product, (±SD) = ±standard deviation, *M* = Mortality mice, IE = Invasion extensity.

**Table 11.** Comparison between experimental (effectivity) and theoretical activity

| Product | Dose (mg/kg) | $E$ (%)[a] | $\Delta P\%$[b] | (NAA)[c] | (AchI)[c] | (GR)[c] | (t)[c] | (H)[c] |
|---|---|---|---|---|---|---|---|---|
| Compound **1** | 45 | 67 | 51.70 | 0.11 | 0.00 | 0.00 | 0.34 | 0.54 |

[a] Anthelmintic activity measured experimentally (see Biological activity and in vivo experiment).
[b] Anthelmintic activity predicted by Eq. 10; $\Delta P\% = [P(\text{Active}) - P(\text{ Inactive})] \times 100$ (see Chemometric method).
[c] Percentage of probability with which the drug is predicted to be anthelmintic having NAA, AChI, GR, t, H as mechanism of action, respectively, using Eqs. 11–15.

proton and is very lipophilic. As can be seen in molecular structure of compound 1 (see Table 10), the H-atom in the α-carbon of the exocyclic double bond and their lipophilic features (Br-atom, nitro-group and aromatic ring), can explain the mode of action hypothesis, because this lipid soluble substance is capable of carrying protons across membrane, therefore, may act to uncouple oxidative phosphorylation preventing the production of the proton gradient across the inner mitochondrial membrane. Table 10 shows the position of the proton that is able to dissociate from compound **1**. However, this theoretical result must be considered only as preliminary screening results and an experimental corroboration is necessary in the future.

## 4. Concluding remarks

Computational—in silico—screening has become a very important tool in the development of drug discovery today; although it is clear that in silico predictive modeling does not represent a panacea for the industry.[33] In virtual (computational) screening studies as well as in molecular diversity analysis of large data set of compounds, many of the traditional QSAR descriptors are not applicable as they apply to congeneric series.[17] In this paper, atom, atom-type and total linear indices were used to obtain a quantitative LDA model that discriminate anthelmintic compounds from inactive ones; and to perform another model to classify the anthelmintic compounds taking into consideration their modes of action. The obtained models were significant of the statistical point of view. So, we can say that TOMOCOMD-CARDD approach is not only able to discriminate between anthelmintic and non-anthelmintic compound, but also to select anthelmintic compounds from a pool of novel structure. Besides, virtual screening of drugs was carried out by us to prove the usefulness of the present approach to discover new anthelmintic compounds from 2D chemical structure databases or combinatorial libraries. Finally, using the developed models, a new lead candidate has been identified as a promising starting point for the rational design of new 2-furylethylenes derivatives with potent anthelmintic activity. Some works in this direction are at the moment in progress and will be published in a forthcoming paper.

## 5. Experimental

### 5.1. Computational methods: TOMOCOMD-CARDD approach

TOMOCOMD is an interactive program for molecular design and bioinformatics research.[18] The program consists of four subprograms, each one dealing with drawing structures (drawing mode) and calculating 2D and 3D molecular descriptors (calculation mode). The modules are named CARDD (Computed-Aided 'Rational' Drug Design), CAMPS (Computed-Aided Modeling in Protein Science), CANAR (Computed-Aided Nucleic Acid Research) and CABPD (Computed-Aided Bio-Polymers Docking). In this paper, we outline salient features concerned with only one of these subprograms: CARDD. This subprogram was developed based on a user-friendly philosophy. That is to say, this computer graphics software shows a great efficiency of interaction with user, without prior knowledge of programming skills (e.g. practicing pharmaceutics and organic chemist, teacher, university student, and so on).

The calculation of total and local (atom and atom-type) linear indices for any organic molecule (or any drug-like compound) was implemented in the TOMOCOMD-CARDD software.[18] The main steps for the application of this method in QSAR/QSPR and drug design can be briefly summarized as follows:

1. Draw the molecular pseudographs for each molecule of the data set, using the software drawing mode. This procedure is carried out by a selection of the active atomic symbols belonging to different groups in the periodic table.
2. Use appropriated atom weights in order to differentiate the molecular atoms. In this work, we used as atomic property the Mulliken electronegativity[28] for each kind of atom.
3. Compute the total and local linear indices of the molecular pseudograph's atom adjacency matrix. They can be carried out in the software calculation mode, from which one can select the atomic properties and the family descriptor previously to calculate the molecular indices. This software generates a table in which the rows correspond to the compounds and columns correspond to the total and local linear indices or any other family of molecular descriptors implemented in this program.
4. Find a QSPR/QSAR equation by using mathematical techniques, such as multilinear regression analysis (MRA), neural networks (NN), linear discrimination analysis (LDA), and so on. That is to say, one can find a quantitative relation between a property $P$ and the linear indices having, for instance, the following appearance,

$$P = a_0 \boldsymbol{f}_0(x) + a_1 \boldsymbol{f}_1(x) + a_2 \boldsymbol{f}_2(x) + \cdots + a_k \boldsymbol{f}_k(x) + \mathrm{c} \quad (16)$$

where $P$ is the measurement of the property, $\boldsymbol{f}_k(x)$ is the $k$th total (atom and atom-type) linear indices, and the $a_k$'s are the coefficients obtained by the linear regression analysis.

5. Test the robustness and predictive power of the QSPR/QSAR equation by using internal and external cross-validation techniques.

Develop a structural interpretation of obtained QSAR/QSPR model using total and local (atom and atom-type) linear indices as molecular descriptors.

The descriptors calculated in this work were the following: (1) $\boldsymbol{f}_k(x)$ are the $k$th total linear indices of the H-suppressed molecular pseudograph, (2) $\boldsymbol{f}_{k\mathrm{L}}(x_{\mathrm{E}})$ are the $k$th local (atom-type = heteroatoms: S, N, O) linear indices of the H-suppressed molecular pseudograph, and (3) $\boldsymbol{f}_{k\mathrm{L}}{}^{\mathrm{H}}(x_{\mathrm{E-H}})$ are the $k$th local (atom-type = H-atoms bonding to heteroatoms: S, N, O) linear indices considering H-atoms in the molecular pseudograph.

### 5.2. General data set

The quality of the classification models relies in a great percentage on the quality of the selected training data set. The most critical aspect of the construction of this series is to guarantee a great molecular structural variability. Here we consider a data set consisting of a great number of molecular entities, some of them reported as anthelmintics[1,3,29,30] and the rest with a series of other pharmacological uses.[29,30]

The data set of active compounds was selected by considering representatives of most of the different structural patterns and action mechanisms of anthelmintic activity. In this sense, we included compounds with the following modes of action: (1) cholinergic agonists of the imidazothiazole type (levamisole and butamisole), tetrahydropyrimidines (pyrantel, morantel, and oxantel), quaternary ammonium salts (bephenium and thenium), and the pyrimidines (methyridine), (2) cholinesterase inhibitors (dichlorvos and haloxon), (3) GABA agonist (piperazine), (4) glutamate-gated chloride receptor potentiators (ivermectin, abamectin, and doramectin), (5) increased calcium permeability (praziquantel), (6) inhibition of microtubule formation, β-tubulin binding (benzimidazoles, albendazole, and netobimin), (7) oxidative phosphorylase uncoplers, often called proton ionophores (salicylanilides and substituted phenols), (8) inhibitors of malate metabolism (diamphenethide), (9) inhibition of phosphoglycerate kinase and mutase (clorsulon), and (10) inhibitors of arachidonic acid metabolism (diethylcarbamzine).[1,3] Other compounds that have not been found or defined a specific mode of action, but had been reported as anthelmintics were also included, such as nitroimidazoles and hicantone derivatives.

The inactive compounds were selected in a random way of a great database, with different pharmacological uses (antibacterial, antifungal, antihypertensive, anticancer, hipnotic/sedative, antidepressant, and so on). The classification of these compounds as 'inactive' (without anthelmintic activity) does not guarantee that some of these compounds present some anthelmintic activity even not detected. The data set was randomly divided into two subsets; one of these was used as training (learning) set and the other one as test set.

### 5.3. Chemometric methods

All statistical analyses were performed with STATISTICA software package.[48] The tolerance parameter (proportion of variance that is unique to the respective variable) used was the default value for minimum acceptable tolerance, which is 0.01. Forward stepwise was fixed as the strategy for variable selection. The principle of parsimony (Occam's razor) was taken into account as strategy for model selection. In this connection, we select the model with higher statistical signification but having as few parameters ($a_k$) as possible. The LDA is used in order to generate the classifier functions on the basis of the simplicity of the method.[17,22,49–51] In order to test the quality of the discriminant function derived we used the Wilks' $\lambda$ ($U$-statistic) and the Mahalanobis distance ($D^2$). The Wilks' $\lambda$ statistical helpful to value the total discrimination and can take values between zero (perfect discrimination) and one (no discrimination). In $D^2$, $D$ indicates the separation of the respective groups. Afterward, the classification trees module was used to carry out the leave-$n$-out cross-validation routine. In addition, to assess the predictive power of the model an external (test) set were used. In this case, the values of 1 and $-1$ were assigned to the active and inactive compounds, respectively. The classification of cases was performed by means of the posterior classification probabilities. By using the models one compound can then be classified as active, if $\Delta P\% > 0$, being $\Delta P\% = [P(\mathrm{Active}) - P(\mathrm{Inactive})] \times 100$ or as inactive otherwise. $P$(Active) and $P$(Inactive) are the probabilities that the equations classify a compound as active and inactive, respectively. The probability that a case belongs to a particular group is basically proportional to the Mahalanobis distance from that group centroid. In closing, the posterior probability is the probability, based on our knowledge of the values of other variables that the respective case belongs to a particular group.

### 5.4. Biological activity and in vivo experiment

We developed an in vivo experiment to measure the chemical effectiveness against *F. hepatica*; in this case, an experimental technique reported in the literature was selected for biological material processing and *F. hepatica* egg extraction.[52] Mitterpak et al.'s technique for host (*Lymnaea cubensis*) invasion was carried out.[53] Afterwards we followed the steps reported by Olazábal et al. to obtain the metacercariae.[54] Metacercariae were conserved in cold until the in vivo experiment.[52]

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.[52,55]

**Table 12.** Experimental design

| Product | Dose (mg/kg) | Safety index |
|---|---|---|
| Compound **1** | 16 | 4 |
| | 45 | 1.42 |

Safety index = maximum tolerated dose/applied dose.

The CBQ organic synthesis laboratory synthesized the compound **1**, with 98% purity. This chemical was tested in order to evaluate their effectiveness against *F. hepatica* according to the following experimental conditions (see also Table 12). Three treatment groups with five mice per group were arranged for the test. One group (infected control group) was treated with Miglyol 810N (administration vehicle). The second group was neither infested nor treated. The remaining group was treated with chemical **1** (see Table 12).

The compound **1** was previously diluted in 10 mL of Miglyol. All mice received 0.2 mL of the compound **1**, for intraperitoneal route. Mouse invasion with metacercariae of *F. hepatica*, two weeks old, 14 days before drug administration, was carried out by Corba et al.'s method.[56] The effectiveness was evaluated based on: (1) Determination of the $E\%$ index. This is a quantitative indicator of effectiveness introduced by Steward[57] and defined as $E\% = [(XC - XT)/XC] \times 100$.[58] 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. The effectiveness was measured based on the elimination or not of *F. hepatica*, in its juvenile stage, as shown by laboratory diagnostics, using the helminthological necropsy on day 7 after the inoculated treatment.[56] (2) Determination of the hepatic index,[59] by use of the formula $A = (B/C) \times 100$. In this case, $A$ = hepatic index, $B$ = liver weight and $C$ = body weight. (3) Degrees of lesions of the liver.[56] (4) Spleen relative weight.[60] (5) Intensity of invasion making use of the formula $I \cdot I = A/B$, where $A$ = total amount of parasites, $B$ = total amount of positives. (6) Extension of invasion by use of the formula $\% E \cdot I = [T(t)/T(a)] \times 100$, where $\%E \cdot I = \%$ of invasion extensity, $T(t)$ = number of total positives, $T(a)$ = total of infected animals.[61] (7) Gain of weight (final weight) – (initial weight).

From these different effectiveness indexes,[59–61] the $E\%$ index was selected. We used this index in spite of the existence of other (more recently defined) effectiveness parameters because is a direct expression of effectiveness that one can compare with $\Delta P\%$.

## Acknowledgements

Authors would like to offer our sincere thanks to the anonymous referees for their critical opinions about the manuscript, which have significantly contributed to improving its presentation and quality. I am also indebted to the journal editor, Dr. Chi-Huey Wong for his kind attention.

## References and notes

1. Martin, R. J.; Robertson, A. P.; Bjorn, H. *Parasitology* **1997**, *114*, S111.
2. Pichard, R. K. *Vet. Parasitol.* **1994**, *54*, 259.
3. Martin, R. J. *Vet. J.* **1997**, *154*, 11.
4. González-Díaz, H.; Olazábal, E.; Castañedo, N.; Hernández, I.; Morales, A.; Serrano, H. S.; González, J.; Ramos, R. *J. Mol. Mod.* **2002**, *8*, 237.
5. Cioli, D. *Parasitol. Today* **1998**, *14*, 418.
6. Conceicao, M. J.; Argento, C. A.; Correa, A. *Memorias instituto Oswaldo Cruz.* **2000**, *95*, 375.
7. Liang, Y. S.; Coles, G. C.; Doenhoff, M. *J. Trop. Med. Int. Health* **2000**, *5*, 72.
8. Reynoldson, J. A.; Behnke, J. M.; Pallant, L. J.; Macnish, M. G.; Gilbert, F.; Giles, S.; Spargo, R. J.; Thompson, R. C. *Acta Trop.* **1997**, *68*, 301.
9. Sangster, N. C.; Gill, J. *Parasitol. Today* **1999**, *15*, 141.
10. Coles, C. C. *J. Med. Microbiol.* **1999**, *48*, 323.
11. Moll, L.; Gaasenbeek, C. P.; Vellena, P.; Borgsteede, F. H. *Vet. Parasitol.* **2000**, *91*, 153.
12. Rothwell, J.; Sangster, N. *Int. J. Parasitol.* **1997**, *27*, 313.
13. Walters, W. P.; Sthal, M. T.; Murcko, M. A. *Drug Discovery Today* **1998**, *3*, 160.
14. Venkatesh, S.; Lipper, R. A. *J. Pharm. Sci.* **2000**, *89*, 145.
15. Hann, M.; Green, R. *Curr. Opin. Chem. Biol.* **1999**, *3*, 379.
16. Estrada, E. *Environ. Res.* **2000**, *11*, 55.
17. Estrada, E.; Peña, A. *Bioorg. Med. Chem.* **2000**, *8*, 2755.
18. Marrero-Ponce, Y.; Romero, V. TOMOCOMD software, Central University of Las Villas, 2002. TOMOCOMD (TOpological MOlecular COMputer Design) for Windows, version 1.0 is a preliminary experimental version; in future a professional version will be obtained upon request to Marrero: yovanimp@qf.uclv.edu.cu or ymarrero77@yahoo.es.
19. Marrero-Ponce, Y. *Molecules* **2003**, *8*, 687.
20. Marrero-Ponce, Y. *J. Chem. Inf. Comput. Sci.* doi:10.1021/ci049950k.
21. Marrero-Ponce, Y.; Cabrera, M. A.; Romero, V.; Ofori, E.; Montero, L. A. *Int. J. Mol. Sci.* **2003**, *4*, 512.
22. Marrero-Ponce, Y.; Cabrera, M. A.; Romero, V.; González, D. H.; Torrens, F. *J. Pharm. Pharm. Sci.* **2004**, *7*, 186.
23. Marrero-Ponce, Y.; Nodarse, D.; González-Díaz, H.; Ramos de Armas, R.; Romero-Zaldivar, V.; Torrens, F.; Castro, E. *Int. J. Mol. Sci.* **2004**, *5*, 276–293 *CPS: physchem/0401004*.
24. Marrero-Ponce, Y.; Castillo-Garit, J. A.; Olazabal, E.; Serrano, H. S.; Morales, A.; Castañedo, N.; Ibarra-Velarde, F.; Huesca-Guillen, A.; Jorge, E.; del Valle, A.; Torrens, F.; Castro, E. A. *J. Comput. Aided Mol. Des.*, in press.
25. Marrero-Ponce, Y.; Iyarreta-Veitía, M.; Montero-Torres, A.; Romero-Zaldivar, C.; Brandt, C. A.; Ávila, P. E.; Kirchgatter, K.; *J. Med. Chem.*, submitted for publication.
26. Marrero-Ponce, Y. *Bioorg. Med. Chem.* doi:10.1016/j.bmc.2004.09.034.
27. Marrero-Ponce, Y.; González-Díaz, H.; Romero-Zaldivar, V.; Torrens, F.; Castro, E. A. *Bioorg. Med. Chem.* **2004**, *12*, 5331.
28. Cotton, F. A.; *Advanced Inorganic Chemistry*. Revolucionaria: Havana, 1966.
29. Negwer, M. *Organic-Chemical Drugs and their Synonyms*; Akademie: Berlin, 1987.
30. The Merck Index, 12th ed.; Chapman & Hall, 1996.
31. Baldi, P.; Brunak, S.; Chauvin, Y.; Andersen, C. A.; Nielsen, H. *Bioinformatics* **2000**, *16*, 412.
32. Golbraikh, A.; Tropsha, A. *J. Mol. Graph. Modell.* **2002**, *20*, 269.
33. Watson, C. *Biosilico* **2003**, *3*, 83.

34. Lajiness, M. S. In *Computational Chemical Graph Theory*; Rouvray, D. H., Ed.; Nova Science: New York, 1990; 299.
35. Walters, W. P.; Stahl, M. T.; Murcko, M. A. *Drug Discovery Today* **1998**, *3*, 160.
36. Walchshofer, N.; Delabre-Defayolle, I.; Paris, J.; Petavy, A. F. *J. Pharm. Sci.* **1990**, *7*, 79.
37. Ames, J. R.; Castognoli, N.; Ryan, M. D.; Kavacic, P. *Free Radical Res. Commun.* **1986**, *2*, 107.
38. Ho, N. F. H.; Sims, S. M.; Vidmar, T. J.; Day, J. S.; Barsuhn, C. L.; Thomas, E. M.; Geary, T. G.; Thompson, D. *J. Pharm. Sci.* **1994**, *83*, 1052.
39. Conder, G.; Kuo, M.; Marshall, V.; Zielinski, R.; Dioxapyrrolomycin as an antiparasitic agent and compositions useful therefore. United States Patent 5, 643, 940, 1997.
40. Scherkenbeck, J. Verwendung von cyclischen depsipeptiden mit 12 Ringatomen zur Bekampfung von Endoparasiten, neue cyclische Depsipeptide mit 12 Ringatomen und Verfahren zu Ihrer Herstellung. European Patent 0 664 297 A1, 1988.
41. Serban, A.; Watson, K.; Wilshire, C.; Forsyth, B. 5-Substituted pyrimidine derivatives active against internal animal parasites. United States Patent 4, 460, 588, 1984.
42. Horrobin, D. F. *J. R. Soc. Med.* **2000**, *93*, 341.
43. Oprea, T. I. *Curr. Opin. Chem. Biol.* **2002**, *6*, 384.
44. Oprea, T. I. *J. Comput.-Aided Mol. Des.* **2002**, *16*, 325.
45. Goodman, A. G.; Goodman, L. S.; Gilman, A. In *Las bases Farmacológicas de la Terapeutica*, 9na ed.; McGraw-Hill: Mexico, 1996; Vol. II.
46. Alberts, B.; Bray, D.; Lewis, J.; Raff, M.; Roberts, K.; Watson, J. D. *Molecular Biology of the Cell*, 3rd ed.; Garland Publishing: New York, London, 1994.
47. Blondeau, J. M.; Castañedo, N.; Gonzalez, O.; Medina, R.; Silveira, E. *Antimicrob. Agents Chemother.* **1999**, *11*, 1663.
48. STATISTICA ver. 5.5, Statsoft, Inc., 1999.
49. González-Díaz, H.; Marrero-Ponce, Y.; Hernández, I.; Bastida, I.; Tenorio, E.; Nasco, O.; Uriarte, E.; Castañedo, N.; Cabrera, M. A.; Aguila, E.; Marrero, O.; Morales, A.; Pérez, M. *Chem. Res. Toxicol.* **2003**, *16*, 1318.
50. Estrada, E.; Peña, A.; Garcia-Domenech, R. *J. Comput. Aided Mol. Des.* **1998**, *2*, 583.
51. Julian-Ortiz, J. V.; Gálvez, J.; Muños-Collado, C.; García-Domenech, R.; Gimeno-Cardona, C. *J. Med. Chem.* **1999**, *42*, 3308.
52. Milligen, F. J.; Cornelissen, J. B.; Guasenbeek, C. P.; Bokhout, B. A. *J. Immunol. Methods* **1998**, *213*, 183.
53. Mitterpak, J.; Mendéz, M.; Mauri, M. *Serie Biol.* **1972**, *30*, 1.
54. Olazábal, E.; Morales, A.; Serrano, H.; Brito, E. *Vet. Méx.* **1999**, *30*, 109.
55. Ping, C.; Hayes, A. *Acute Toxicity and Eyes Irritancy: Principles and Methods of Toxicology*, 3rd ed.; Raven Press Ltd: New York, 1994.
56. Corba, J.; Velebny, S.; Spaldonova, R. *Helminthology* **1981**, *18*, 43.
57. Steward, J. S. *Parasitology* **1955**, *45*, 231.
58. Wood, I. B.; Amaral, N. K.; Bairden, K.; Duncan, J.; Kassai, T.; Malone, J. B.; Pankavich, J. A.; Reinecke, R. K.; Slocombe, O.; Taylor, S. M.; Vercruysse, J. *Vet. Parasitol.* **1995**, *58*, 181.
59. Tsocheva, N.; Krustev, L.; Polyakova, O. *Helminthology* **1992**, *27*, 261.
60. Poitou, I.; Baeza, E.; Boulard, C. *Vet. Parasitol.* **1992**, *45*, 59.
61. Espaine, L.; Lines, R.; Demedio, J. Manual de Parasitología y enfermedades Parasitarias I. Editorial MES, Habana, 1996; p 103.