Available online at www.sciencedirect.com

SCIENCE DIRECT

Bioorganic & Medicinal Chemistry

Bioorganic & Medicinal Chemistry 13 (2005) 6264–6275

A novel non-stochastic quadratic fingerprints-based approach for the 'in silico' discovery of new antitrypanosomal compounds

Alina Montero-Torres,[a,*] María Celeste Vega,[b] Yovani Marrero-Ponce,[a,c] Miriam Rolón,[b] Alicia Gómez-Barrio,[b] José Antonio Escario,[b] Vicente J. Arán,[d] Antonio R. Martínez-Fernández[b] and Alfredo Meneses-Marcel[e]

[a]*Department of Synthesis and Drug Design, Chemical Bioactive Center, Central University of Las Villas, Santa Clara 54830, Villa Clara, Cuba*

[b]*Department of Parasitology, Faculty of Pharmacy, UCM, 28040 Madrid, Spain*

[c]*Department of Pharmacy, Faculty of Chemistry and Pharmacy. Central University of Las Villas, Santa Clara, 54830, Villa Clara, Cuba*

[d]*Institute of Medicinal Chemistry, CSIC, 28006 Madrid, Spain*

[e]*Department of Parasitology, Chemical Bioactive Center, Central University of Las Villas, Santa Clara 54830, Villa Clara, Cuba*

Received 28 March 2005; revised 24 June 2005; accepted 24 June 2005
Available online 22 August 2005

Abstract—A non-stochastic quadratic fingerprints-based approach is introduced to classify and design, in a rational way, new antitrypanosomal compounds. A data set of 153 organic chemicals, 62 with antitrypanosomal activity and 91 having other clinical uses, was processed by a *k*-means cluster analysis to design training and predicting data sets. Afterwards, a linear classification function was derived allowing the discrimination between active and inactive compounds. The model classifies correctly more than 93% of chemicals in both training and external prediction groups. The predictability of this discriminant function was also assessed by a leave-group-out experiment, in which 10% of the compounds were removed at random at each time and their activity predicted a posteriori. In addition, a comparison with models generated using four well-known families of 2D molecular descriptors was carried out. As an experiment of virtual lead generation, the present *TOMOCOMD* approach was finally satisfactorily applied on the virtual evaluation of 10 already synthesized compounds. The in vitro antitrypanosomal activity of this series against epimastigotes forms of *Trypanosomal cruzi* was assayed. The model was able to predict correctly the behaviour of these compounds in 90% of the cases.

1. Introduction

Once an almost exclusively rural disease in Latin America, Chagas' disease is now undergoing a change in its epidemiological profile due to rising levels of urbanization and migration. The latest data from the WHO indicate that over 24 million people are infected, or are at least serologically positive for *Trypanosoma cruzi*, which is the causative agent of such infection. This quantity roughly represents 8% of the total Latin American population. Another factor, blood transfusion, is considered the second most frequent route of transmission in endemic countries, given that parasites may survive in whole blood stored for more than 21 days at 4 °C and detection techniques are not always strictly applied.[1]

Medication for Chagas' disease is usually effective when it is given during the acute stage of infection. No medication has been proven to be effective once the disease has progressed to later stages. Moreover, synthetic drugs such as nifurtimox and benznidazole have severe side effects, including cardiac and/or renal toxicity. This explains the need for discovering new effective chemotherapeutic and chemoprophylactic agents against *T. cruzi*.[2,3] In this sense, medicinal chemists need to find new effective drugs in a fast and inexpensive way. In the last few decades, computer-aided drug design approaches have emerged as promising tools for solving this problem.[4–10] With the use of such design strategies,

Keywords: Antitrypanosomal compounds; Chagas' disease; LDA-based-QSAR-model; Non-stochastic quadratic indices; QSAR; *TOMOCOMD* software.

* Corresponding author. Tel.: +53 42 281192/281473; fax: +53 42 281130/281455; e-mail addresses: amontero@uclv.edu.cu; alinamontero@gmail.com

0968-0896/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmc.2005.06.049

the handling and screening of large databases to find reduced sets of potential new drug candidates is possible.[11,12] Thus, the development of computational approaches based on discrimination functions plays an important role, allowing the identification from large chemical libraries of structural subsystems responsible for a property or biological activity, and in this way, the classification of active compounds from inactive ones.

In this context, our research group has recently developed a novel scheme to generate molecular fingerprints based on the application of discrete mathematics and linear algebra theory. The approach (known as *TOMOCOMD* acronym of *To*pological *Mo*lecular *Com*puter *D*esign),[13–16] allows us to perform rational in silico molecular design (selection/identification) and QSAR/QSPR studies. In this sense, this scheme has been applied to the prediction of several physical, physicochemical, chemical, pharmacokinetical as well as biological properties.[17–19] It was, for instance, successfully used in the virtual screening of novel antihelminthic compounds, which were then synthesized and evaluated in vivo on *Fasciola hepatica*.[20,21] Other studies for the rational discovery of novel paramphistomicides,[22] antimalarial [23] and antibacterial [24] compounds were also conducted with the *TOMOCOMD* approach. This method has been extended to consider three-dimensional (3D) features of small/medium-sized molecules on the basis of the application of a trigonometric 3D-chirality correction factor.[25]

In the present study, *TOMOCOMD* strategy is used to find a classification model that allows the discrimination of antitrypanosomal compounds from inactive ones. It is also an objective of the present work to assess the model robustness and predictive power by using external and internal cross-validation techniques. The in silico evaluation of 10 new heterocyclic compounds is finally performed, and their in vitro antitrypanosomal activity against epimastigotes forms of *T. cruzi* investigated. The results of the current study are presented as a starting point for the development of new inexpensive antitrypanosomals.

2. Results and discussion

2.1. Computing non-stochastic quadratic molecular fingerprints

To obtain quantitative structure–property or structure–activity relationships (abbreviated QSPR and QSAR, respectively), it is necessary to convert the molecular structures into numbers that could be later processed statistically; this means that a structural parameterization is required. Such a problem is overcome by means of the computation of molecular descriptors.[26] In the last few decades, a great number of molecular fingerprints have been presented in the literature.[27,28] Atomic, atom-type and total non-stochastic quadratic indices have shown a great ability to encode chemical information, which can be used for the development of QSARs. The theoretical scaffold of this *TOMOCOMD*'s molecular fingerprints has been presented in detail in earlier papers.[14–25] Here just a short overview will be given.

Atomic, atom-type and total molecular quadratic indices have been defined in analogy to the quadratic mathematical maps.[14,16] After constructing the molecular pseudograph's atom adjacency matrix **M**(G) and the molecular vector (**X**), whose components $x_1, \ldots, x_n$ are numeric values or weights (atom-labels or atom-properties) for the vertices of the pseudograph, kth total quadratic indices, $\boldsymbol{q}_k(x)$, can be computed for a given molecule composed of n atoms as shown in Eq. 1,

$$q_k(x) = \sum_{i=1}^{n} \sum_{j=1}^{n} {}^k a_{ij} x_i x_j \qquad (1)$$

where, ${}^k a_{ij}$ are the elements of the kth power of the symmetrical square matrix **M**(G) of the molecular pseudograph (G), and are defined as follows:

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

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

Eq. 1 for $q_k(x)$ can also be written as a single matrix equation:

$$\mathbf{q}_k(\mathbf{x}) = \mathbf{X}^t \mathbf{M}^k \mathbf{X} \qquad (3)$$

where **X** is a column vector (a $n \times 1$ matrix), $\mathbf{X}^t$ the transpose of **X** (a $1 \times n$ matrix) and $\mathbf{M}^k$ the kth power of the matrix **M** of the molecular pseudograph G.

In a similar way, local fragment (atomic and atom-type) formalisms can be developed. The local quadratic indices, $\boldsymbol{q}_{kL}(x)$[14,16] for a fragment containing m atoms can be computed as follows:

$$q_{kL}(x) = \sum_{i=1}^{m} \sum_{j=1}^{m} {}^k a_{ijL} x_i x_j \qquad (4)$$

where ${}^k \boldsymbol{a}_{ijL}$ is the element of the row 'i' and column 'j' of the matrix $\mathbf{M}_L^k$ and is defined as follows:

$$\begin{aligned} {}^k a_{ijL} &= {}^k a_{ij} \text{ if both } v_i \text{ and } v^j \text{ are atoms contained within the molecular fragment} \\ &= 1/2\,{}^k a_{ij} \text{ if } v_i \text{ or } v_j \text{ is an atom contained within the molecular fragment but not both} \\ &= 0 \text{ otherwise} \end{aligned}$$

These local analogues can also be expressed in matrix form:

$$\boldsymbol{q}_{kL}(x) = \mathbf{X}^t \mathbf{M}_L^k \mathbf{X} \qquad (6)$$

For every partition of a molecule into Z molecular fragments there will be Z local molecular fragment matrices $\mathbf{M}_L^k$. The kth power of the matrix **M** is exactly the sum of the kth power of the local Z matrices and in this way, the total quadratic indices are the sum of the quadratic indices of the Z molecular fragments:

$$\boldsymbol{q}_k(x) = \sum_{L=1}^{Z} q_{kL}(x) \tag{7}$$

Atom and atom-type quadratic fingerprints are specific cases of local quadratic indices. In the atom-type quadratic indices formalism, each atom in the molecule is classified into an atom-type (fragment), such as heteroatoms, hydrogen bonding (H-bonding) to heteroatoms (O, N and S), halogen atoms, etc. It has been proved that for all data sets, considering those with a common molecular scaffold as well as those with diverse ones, the kth atom-type quadratic indices contain important structural information.

In the current work, the kth total quadratic indices [$\boldsymbol{q}_k(x)$ and $\boldsymbol{q}_k^{\mathrm{H}}(x)$] and the kth local ones (atom-type = heteroatoms: S, N, O) [$\boldsymbol{q}_{k\mathrm{L}}(x_{\mathrm{E}})$ and $\boldsymbol{q}_{k\mathrm{L}}^{\mathrm{H}}(x_{\mathrm{E}})$] without and with consideration of H-atoms, respectively, were computed.

2.2. Training and test sets design

To obtain mathematical expressions capable of discriminating between active and inactive compounds, the chemical information contained in a great number of compounds with and without the desired biological activity must be statistically processed. Taking into account that the most critical aspect in the construction of a training data set is the molecular diversity of the included compounds, we selected a group of 153 organic chemicals having as much structural variability as possible. The antitrypanosomals considered in this study are representatives of families with diverse structural patterns.[29–38] Figure 1 shows the whole active set collected from the literature for this work.

The selected inactive group included antivirals, sedative/hypnotics, diuretics, anticonvulsivants, hemostatics, oral hypoglycemics, antihypertensives, antihelminthics, anticancer compounds as well as some other kinds of drugs, guaranteeing at the same time a great structural variability.[39]

To split the whole group into two data sets (training and predicting ones), two k-MCA[40,41] were performed for antitrypanosomal and inactive compounds. In this sense, a partition of either active or inactive series of chemicals in several statistically representative classes of compounds is performed. This process ensures that any chemical class identified by the k-MCA will be represented in both, training and test sets.

A first k-MCA (I) split antitrypanosomals into six clusters with 12, 3, 2, 13, 13, and 19 members. The inactive compound series was also partitioned by a second k-MCA (II) into six clusters with 17, 12, 14, 19, 18, and 11 compounds in each case.

Afterwards, the selection of the training and prediction sets was performed by taking compounds belonging to each cluster at random. From these 153 chemicals, 101 were chosen to form the training set, of which 40 were actives and 61 were inactive ones. The remaining group, consisting of 22 antitrypanosomals and 30 compounds with other different biological properties, was prepared as a test set for the external model validation process. These 52 compounds were not used in the development of the classification model. Figure 2 graphically illustrates the above-described procedure.

An inspection of the standard deviation between and within each cluster, the Fisher ratios and the p-levels of significance for each variable, permits us to ensure that the data partition into the respective clusters can be considered as a statistically acceptable process. The kth total and atom-type non-stochastic quadratic indices were used in this analysis, with all variables showing p-levels <0.05 for the Fisher test. The main results are depicted in Table 1.

In this sense, it can be concluded that the data set of antitrypanosomal compounds considered in this study encompasses compounds of six general structural patterns codified by *TOMOCOMD* descriptors and recognized by a k-means cluster analysis.

2.3. Developing a discriminant function

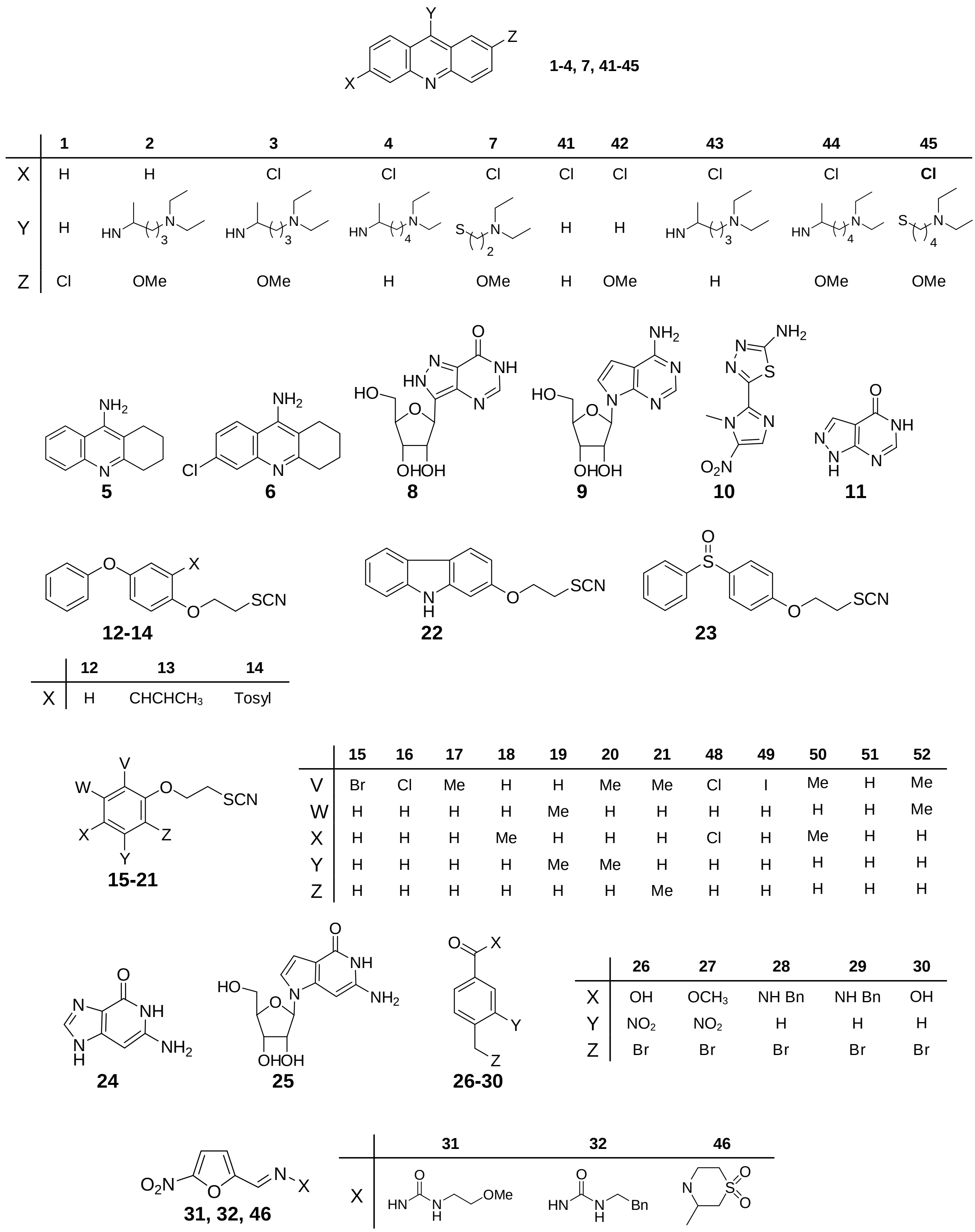
Linear discriminant analysis (LDA) has become an important tool for the prediction of chemicals properties. On the basis of the simplicity of this method many useful discriminant models have been developed and presented by different authors in the literature.[7–10,20–24] As LDA was also the technique used in the generation of a discriminant function in the current work. The principle of parsimony (Occam's razor) was taken into account as the strategy for model selection.[42] This means that the model with higher statistical signification but having as few parameters (a_k) as possible, is selected.

Making use of the LDA technique implemented in the STATISTICA software,[43] the following linear model was obtained:

$$\begin{aligned}\text{Class} = {} & -5.18 + 2.36\times10^{-4}\boldsymbol{q}_7(x) - 1.30\\ & \times10^{-4}\boldsymbol{q}_8(x) + 2.08\times10^{-5}\boldsymbol{q}_9(x) + 0.97\\ & \times10^{-7}\boldsymbol{q}_{14\mathrm{L}}(x_{\mathrm{E}}) - 2.92\times10^{-8}\boldsymbol{q}_{15\mathrm{L}}(x_{\mathrm{E}})\\ & - 3.28\times10^{-4}\boldsymbol{q}_4^{\mathrm{H}}(x)\end{aligned} \tag{8}$$

$N = 101$, $\lambda = 0.36$, $D^2 = 7.10$, $F(6,94) = 27.16$, $p < 0.0001$. where, N is the number of compounds, λ is Wilk's coefficient, F is the Fisher ratio, D^2 is the squared Mahalanobis distance and p-value is the significance level. The antitrypanosomal activity was codified by a dummy variable 'Class', which indicates either the presence of an active compound (Class = 1) or an inactive one (Class = −1). The classification of cases was performed by means of the posterior classification probabilities, which is the probability that the respective case belongs to a particular group (active or inactive). By using the models, each compound can be then either classified as active, if $\Delta P\% > 0$, where $\Delta P\% = [P(\text{active}) - P(\text{inactive})] \times 100$, or as inactive otherwise. Compounds with $\Delta P\% < 5\%$ were considered as non-classified. Table 2 shows these results.

As can be computed from the results showed in Table 2, model 8 correctly classified 93.02% of the whole training data set (accuracy). This model showed a high Mathews' correlation coefficient (MCC) of 0.87. MCC is a measure that may provide a much more balanced evaluation of the prediction than the percentages of good



	1	2	3	4	7	41	42	43	44	45
X	H	H	Cl	Cl	Cl	Cl	Cl	Cl	Cl	Cl
Y	H					H	H			
Z	Cl	OMe	OMe	H	OMe	H	OMe	H	OMe	OMe

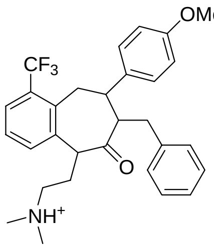
	12	13	14
X	H	$CHCHCH_3$	Tosyl

	15	16	17	18	19	20	21	48	49	50	51	52
V	Br	Cl	Me	H	H	Me	Me	Cl	I	Me	H	Me
W	H	H	H	H	Me	H	H	H	H	H	H	Me
X	H	H	H	Me	H	H	H	Cl	H	Me	H	H
Y	H	H	H	H	Me	Me	H	H	H	H	H	H
Z	H	H	H	H	H	H	Me	H	H	H	H	H

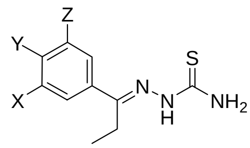
	26	27	28	29	30
X	OH	OCH_3	NH Bn	NH Bn	OH
Y	NO_2	NO_2	H	H	H
Z	Br	Br	Br	Br	Br

	31	32	46
X			

Figure 1. Structures of active compounds in training and test groups.

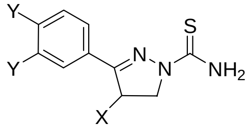


33

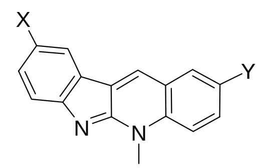


34, 35, 58

	34	35	58
X	Br	CF_3	Cl
Y	H	H	Cl
Z	H	CF_3	H



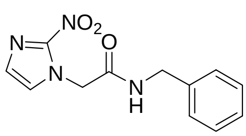
37, 38, 59-61



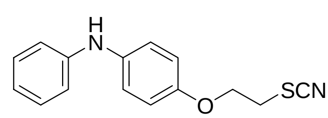
38-40, 62

	36	37	59	60	61
X	Me	Me	Me	H	H
Y	Cl	CF_3	Br	Cl	CF_3
Z	Cl	H	H	Cl	H

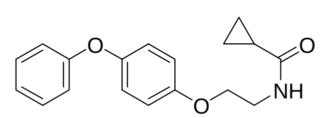
	38	39	40	62
X	H	H	H	CN
Y	OMel	H	NO_2	OMe



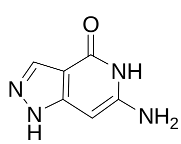
47



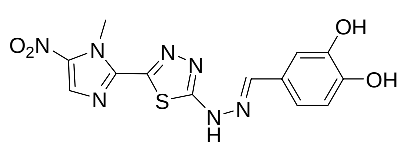
53



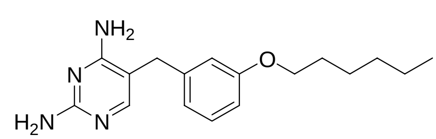
54



55



56



57

Figure 1 (*continued*)

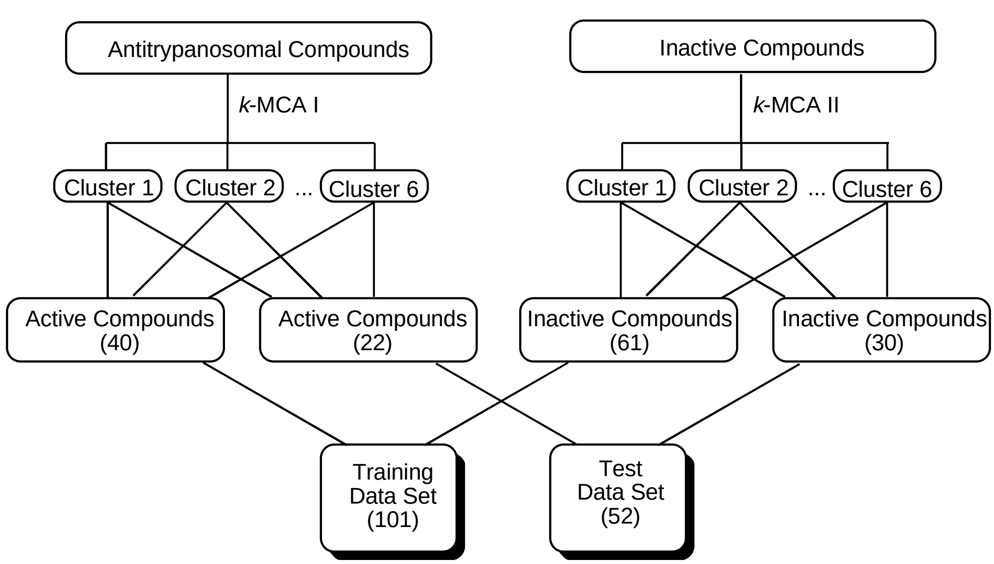


Figure 2. Training and test data sets design throughout *k*-means cluster analysis.

Table 1. Main results of the *k*-means cluster analysis, for antitrypanosomal and inactive compounds

	Analysis of variance			
Total and atom-type quadratic indices	Between SS[a]	Within SS[b]	Fisher ratio (F)	p-level[c]
Antitrypanosomal agents clusters (k-MCA I)				
$\boldsymbol{q}_6(x)$	29.27	3.03	108.10	0.00
$\boldsymbol{q}_7(x)$	32.78	3.25	112.91	0.00
$\boldsymbol{q}_8(x)$	35.06	3.34	117.24	0.00
$\boldsymbol{q}_9(x)$	38.80	4.01	108.24	0.00
$\boldsymbol{q}_{10}(x)$	41.12	4.37	105.37	0.00
$\boldsymbol{q}_{13\mathrm{L}}(x_\mathrm{E})$	21.07	4.90	48.10	0.00
$\boldsymbol{q}_{14\mathrm{L}}(x_\mathrm{E})$	19.53	4.81	45.38	0.00
$\boldsymbol{q}_{15\mathrm{L}}(x_\mathrm{E})$	21.38	4.60	51.96	0.00
$\boldsymbol{q}_4^\mathrm{H}(x)$	27.35	6.92	44.23	0.00
$\boldsymbol{q}_8^\mathrm{H}(x)$	34.81	3.82	102.06	0.00
$\boldsymbol{q}_9^\mathrm{H}(x)$	37.28	3.60	115.76	0.00
$\boldsymbol{q}_{10}^\mathrm{H}(x)$	39.06	3.47	126.07	0.00
$\boldsymbol{q}_{11}^\mathrm{H}(x)$	41.68	3.52	132.52	0.00
$\boldsymbol{q}_{12}^\mathrm{H}(x)$	43.61	3.58	136.16	0.00
Non-antitrypanosomal agents clusters (k-MCA II)				
$\boldsymbol{q}_6(x)$	60.14	9.33	109.57	0.00
$\boldsymbol{q}_7(x)$	53.57	9.60	94.82	0.00
$\boldsymbol{q}_8(x)$	56.86	9.70	99.64	0.00
$\boldsymbol{q}_9(x)$	52.16	10.34	85.73	0.00
$\boldsymbol{q}_{10}(x)$	55.78	10.44	90.76	0.00
$\boldsymbol{q}_{13\mathrm{L}}(x_\mathrm{E})$	121.32	11.01	187.23	0.00
$\boldsymbol{q}_{14\mathrm{L}}(x_\mathrm{E})$	126.71	10.94	196.72	0.00
$\boldsymbol{q}_{15\mathrm{L}}(x_\mathrm{E})$	123.22	14.02	149.32	0.00
$\boldsymbol{q}_4^\mathrm{H}(x)$	77.20	16.76	78.27	0.00
$\boldsymbol{q}_8^\mathrm{H}(x)$	58.04	11.25	87.66	0.00
$\boldsymbol{q}_9^\mathrm{H}(x)$	53.15	11.47	78.74	0.00
$\boldsymbol{q}_{10}^\mathrm{H}(x)$	52.98	11.46	78.57	0.00
$\boldsymbol{q}_{11}^\mathrm{H}(x)$	49.55	11.96	70.39	0.00
$\boldsymbol{q}_{12}^\mathrm{H}(x)$	50.01	12.18	69.75	0.00

[a] Variability between groups.
[b] Variability within groups.
[c] Level of significance.

classification, because it uses all four numbers (true positive, true negative, false positive and false negatives).[44] In addition, the probability of correctly predicting a positive example (sensitivity or hit rate) and the probability that a positive prediction will be correct (specificity) were computed for the model. In both cases, 92.31% was the value obtained. While these two latter measures provide some information on the predictivity for positive observations, the negative predictive value (sensitivity of the negative category) gives a criterion of good classification for the inactive group. In this case, a value of 95.08% was observed.[44] These results, as well as the false positive rate (false alarm rate) are given in Table 3.

Every statistical model that is generated based on a previously selected data set of observations, includes information of just a portion of the universe and has an error range, which the researcher tries to minimize during the modeling process. In this sense, the false positive rate, as well as the false negative rate, are used as measures of the error range and the confiability of the model. A correct selection of a training data sets can reduce the magnitude of both measures. We took this aspect into consideration and built a training data set choosing chemicals with so much structural variability as possible. Despite the previous precaution, it can so happen that the combinations of some structural patterns of a positive case, for instance, results in mathematical values, which are closer to those obtained from the combination of structural fragments in a negative observation. In such a case the model will not recognize the true class of the observation. In the present study, three active and three inactive compounds were mis-classified. Here it is also important to note that the declaration of each non-antitrypanosomal compound as inactive does not mean that antitrypanosomal side effects do not exist. This can include organic drugs for which antitrypanosomal activity has been left undetected so far. In this sense, any discriminant model can be continuously transformed and improved, taking into consideration unavailable information at the time of the model's development. This problem can affect, to some degree, the results of further classification. It is possible to ensure the absence of antitrypanosomal effects just by testing the biological activity. In this sense, we recommend carrying out the biological assays for previously declared inactive compounds, for which the model gives a positive classification.

Considering that a discriminant model could be accepted or rejected depending on its predictive power, it is clear to see that validation processes constitute obligatory steps for the assessment of any structure–activity relationship. As Golbraikh and Tropsha emphasized, the

Table 2. Classification of active and inactive compounds included in the training set using Model 8

Compound	$\Delta P\%$[a]	Class	Compound	$\Delta P\%$[a]	Class
Training active group					
1	0.99	+	21	0.80	+
2	0.98	+	22	1.00	+
3 (Mecaprine)	0.99	+	23	1.00	+
4	0.98	+	24	0.43	+
5	0.87	+	25	0.84	+
6 (Chlorotacrine)	0.94	+	26	0.42	+
7	1.00	+	27	0.50	+
8 (Formycin B)	0.89	+	28	0.03	NC
9 (Tubercidin)	0.68	+	29	0.62	+
10 (Megazol)	0.95	+	30	−0.73	–
11 (Allopurinol)	0.93	+	31	−0.27	–
12	0.98	+	32	0.53	+
13	1.00	+	33	0.99	+
14	0.76	+	34	−0.31	–
15	0.77	+	35	0.63	+
16	0.78	+	36	0.68	+
17	0.65	+	37	0.52	+
18	0.60	+	38	1.00	+
19	0.75	+	39	1.00	+
20	0.77	+	40	1.00	+
Training inactive group					
3-Episiostatin B	−1.00	–	Ganglefene	−0.92	–
Thiacetazone	−0.68	–	Metadiphenil bromidum	−0.68	–
TBHQ	−0.74	–	Quateron	−0.94	–
Cloral betaine	−0.99	–	Pancuronium	−1.00	–
Vernelan	−0.99	–	Ethylene	−0.99	–
Cetohexazine	−0.81	–	Dioxychlorane	−0.99	–
Carbavin	−0.97	–	Aliflurane	−0.90	–
Phenacemide	−0.97	–	Vinyl ether	−0.99	–
Tetharbital	−0.37	–	Tiouracilo	−0.55	–
Brofoxine	−0.46	–	Thiamazol methyl iodide	−0.86	–
Norantoin	−0.54	–	Diclofutime mesilate	0.95	+
Orotonsan Fe	−0.69	–	Percloroetane	−0.98	–
Ferrocholinate	−1.00	–	Lindane	−0.99	–
Ferrosi ascorbas	−0.74	–	Nitrodan	0.92	+
Arecoline	−0.96	–	Ascaridole	−1.00	–
Butanolum	−1.00	–	Pyrantel tartrate	−0.80	–
Etamsylate	−0.99	–	Fentanilo	−0.81	–
Sango-Stop	−0.99	–	Tenalidine tartrate	−0.98	–
Besunide	−0.96	–	Dioxoprometazine	−1.00	–
Spironolactone	−0.97	–	*N*-hidroxymetil-*N*-metilurea	−0.99	–
Glycerol	−0.99	–	2,4,5-triclofenol	−0.19	–
Propamin"soviet	−0.99	–	Norgamem	−0.99	–
Cystamine	−1.00	–	Furtrethonium iodide	−0.99	–
Amifostine	−1.00	–	Isofenefrine	−0.93	–
Adeturon	−1.00	–	Phenylethanolamine	−0.96	–
Glisolamide	−0.38	–	Cefalexin	0.54	+
Glibutimine	−0.81	–	Streptomycin	−0.99	–
Ag 307	−0.93	–	Azirinomycin	−0.95	–
Bromcholine	−1.00	–	Gentamicin A1	−1.00	–
Mebetide	−0.99	–	(2-Hidroxypropyl trimetilamonium hydroxide)	−1.00	–
Minoxidil	−0.78	–			

[a] Results of the classification of compounds obtained from Eq. 8 (using non-stochastic quadratic indices): $\Delta P\% = [P(\text{active}) - P(\text{inactive})] \times 100$, NC = not classified.

Table 3. Overall measures of accuracy obtained in the training and prediction sets for the model 8

	Matthews corr. coefficient	Accuracy (%)	Sensitivity (hit rate%)	Specificity (%)	False alarm rate (%)	Predictive value (–) (%)
Training set	0.87	93.06	92.31	92.31	4.92	95.08
Test set	0.88	94.23	90.91	95.24	3.33	96.67

predictive ability of a QSAR model can only be estimated using an external test set of compounds which were never used for the development of the model.[45] In this sense, it is important to ensure, that the prediction algorithm is able to perform well on novel data from the same data domain. In our case, an external prediction data set was evaluated as the first validation experiment. The computation of some performance measures such as Matthews' correlation coefficient, percentage of global good classification (accuracy), sensitivity, specificity, false alarm rate and negative predictive value (sensitivity of the negative category) permitted us to carry out the assessment of the model. The results for this validation process are given in Table 3.

The classification's results using model 8 for active and inactive compounds in the selected test set are shown in Table 4.

A second validation experiment was also developed on the basis of a leave-group-out internal cross-validation strategy.[46] In this case, groups of compounds including 10% of the training data set are left out and predicted later for the model obtained with the remaining 90%. This process was repeated 10 times for each one of the 10 unique subsets selected at random and each observation predicted once (in its group of left-out observations). The overall mean for this process (10% full leave-out cross-validation) was used as a good indication of robustness and stability of the obtained model. In Table 5, the results of classification for each 10%-group, as well as the classification for the remaining training set leaving out each one of those groups are presented. From these results we can conclude that this experiment also shows that our model had a robust and stable behaviour.

No previous reports related to the application of pattern recognition techniques to the selection of antitrypanosomal compounds from a heterogeneous series of compounds were found in the literature. In this sense, the

Table 5. Predictability based on the use of 10 randomly selected subsets (LGO cross-validation) of the LDA Model

	% global good classification	
	Test set (10%)	Remaining training set
Group		
1	100.00	93.40
2	100.00	93.40
3	100.00	93.40
4	100.00	93.40
5	80.00	92.30
6	90.00	93.40
7	90.00	94.50
8	80.00	95.60
9	90.00	92.30
10	90.90	93.33
Overall mean	91.38	93.50
Standard deviation (%)	7.86	0.96

Table 4. Classification of active and inactive compounds included in test series using the model 8

Compound	$\Delta P\%$[a]	Class	Compound	$\Delta P\%$[a]	Class
Test active set					
41	1.00	+	52	0.79	+
42	1.00	+	53	0.97	+
43	0.98	+	54	−0.07	–
44	0.99	+	55	0.99	+
45	0.99	+	56 (Brazilizone A)	1.00	+
46 (Nifurtimox)	−0.95	–	57	0.58	+
47 (Benznidazol)	0.42	+	58	0.22	+
48	0.91	+	59	0.29	+
49	0.76	+	60	0.59	+
50	0.77	+	61	0.41	+
51	0.43	+	62	1.00	+
Test inactive set					
Amantadine	−1.00	–	Cyclopropane	−0.99	–
Mizoribine	−0.16	–	Basedol	−0.95	–
Triclofos	−1.00	–	Mipimazole	−0.99	–
Nitroinosite	−0.92	–	Didym levulinate	−1.00	–
Methenamine	−0.99	–	Metriponate	−1.00	–
Cobalti glutamas	−1.00	–	Prasterone	−0.98	–
Cobalti besilas	−1.00	–	Febensamin	−1.00	–
Canrenone	−0.92	–	Guanazole	−0.99	–
Urea	−1.00	–	Fluorembichin	−1.00	–
Pallirad	−0.99	–	Mitoguazone	−0.99	–
Quimbosan	−0.99	–	Acetylcholine	−1.00	–
Glicondamide	0.65	+	Methacholine chloride	−1.00	–
RMI 11894	−1.00	–	Dopamine	−0.78	–
Barbismetylii iodidum	−0.59	–	Ampicillin	−0.42	–
Frigen 113	−0.98	–	Kanamycin A	−1.00	–

[a] Results of the classification of compounds obtained from Eq. 8 (using non-stochastic quadratic indices): $\Delta P\% = [P(\text{active}) - P(\text{inactive})] \times 100$, NC = not classified.

present algorithm constitutes a step forward in the search of efficient ways to discover new antitrypanosomal drugs.

With the aim of evaluating the applicability of the present *TOMOCOMD* methodology, four families of 2D molecular descriptors were computed with the DRAGON software[47] and the respective models were generated. The statistical parameters are shown in Table 6. As can be seen, in no case are the classification results better than those obtained using model 8. These results are a proof of the usefulness of the *TOMOCOMD* strategy in the study of this biological property.

2.4. An experiment of rational search of novel antitrypanosomal compounds

The importance and usefulness of QSAR models can only be assessed, first by predicting the activity of new compounds not used in the process of constructing the classification algorithm, and later by, carrying out the biological corroboration of such predictions. With the aim of testing the ability of our model to detect new lead compounds, we design a simulated virtual screening experiment using model 8. To avoid the manipulation of large databases of chemicals, and just as an example of applicability of our approach, we selected a series of 10 newly synthesized heterocyclic compounds obtained by one of our research groups.[48–51] As first step of this virtual screening, all structures were drawn using the drawing mode implemented in the *TOMOCOMD* software. Next, the *k*th total quadratic indices [$\boldsymbol{q}_k(x)$ and $\boldsymbol{q}_k^{\mathrm{H}}(x)$] and the *k*th local ones (atom-type = heteroatoms: S, N, O) [$\boldsymbol{q}_{k\mathrm{L}}(x_{\mathrm{E}})$ and $\boldsymbol{q}_{k\mathrm{L}}^{\mathrm{H}}(x_{\mathrm{E}})$] without and with consideration of H-atoms, respectively, were computed. Each compound was evaluated using model 8 and finally assayed in vitro against epimastigotes forms of *T. cruzi*. Epimastigotes are the extracellular multiplying forms of the mentioned parasite which are relatively sensitive to a drug action. They can be easily cultured and are therefore, an excellent platform for preliminary in vitro screening of antitrypanosomal activity.[52] However, once this first assays have been performed, more selective methods are required to determine the activity of novel compounds. In the current work, the preliminary screening on epimastigotes forms was the election way to evaluate the activity of the predicted compounds, whose structures are shown in Figure 3.

The results of the prediction process using model 8, as well as the anti-epimastigotes percentage for each assayed compound, are summarized in Table 7. In this case, nifurtimox was employed as the reference drug.

As predicted, five compounds (compounds **1s–5s**) showed trypanocidal activity (%AE > 70). Compounds **2s**, **3s** and **5s** are only active against epimatigotes at

Table 6. Comparison between model 8 and four models obtained using different kinds of 2D descriptors

Models	Matthews corr. coefficient		Accuracy (%)		Number of variables
	Training set	Test set	Training set	Test set	
Model 8	0.87	0.88	93.06	94.23	6
Topological descriptors	0.80	0.68	85.14	84.61	6
Molecular walk counts	0.45	0.42	69.31	69.23	4
BCUT descriptors	0.64	0.60	81.19	78.85	6
2D autocorrelations	0.86	0.85	91.09	92.31	6

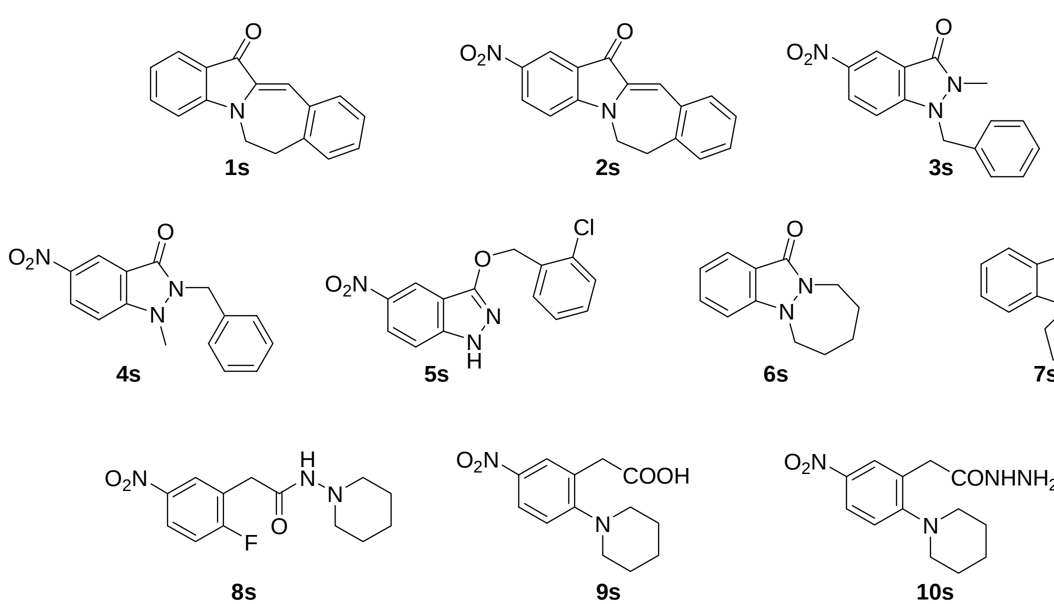


Figure 3. Structures of 10 synthesized compounds evaluated using model 8.

Table 7. Compounds which were evaluated in the present study, their classification ($\Delta P\%$) according to the *TOMOCOMD* approach, their antitrypanosomal activity at three different concentrations (100, 10, and 1 μg/ml) and antitrypanosomal activity of nifurtimox (reference)

Compound	$\Delta P\%$[a]	Class[b]	Obsd[c]	%AE (SS)[d]		
				100(μg/ml)	10(μg/ml)	1(μg/ml)
1s	1.00	+	+	79.90 (1.60)	73.50 (0.70)	29.30 (0.80)
2s	1.00	+	+	99.68 (0.15)	36.78 (0.28)	36.62 (3.11)
3s	0.97	+	+	83.87 (0.40)	23.13 (1.56)	16.41 (0.74)
4s	0.97	+	+	89.60 (0.61)	84.60 (1.50)	82.80 (0.50)
5s	0.99	+	+	100.00 (0.60)	56.82 (0.25)	10.94 (1.70)
6s	−0.04	NC	−	49.78 (0.40)	35.84 (0.50)	29.21 (0.64)
7s	−0.34	−	−	0.00 (1.16)	0.00 (0.84)	0.00 (1.63)
8s	−0.26	−	−	6.90 (4.19)	0.00 (4.22)	0.00 (6.01)
9s	−0.28	−	−	37.56 (1.11)	20.10 (0.40)	9.56 (2.61)
10s	−0.10	−	−	58.65 (3.70)	10.11 (4.98)	4.38 (2.42)
Nifurtimox				98.73 (0.56)	90.05 (1.80)	75.50 (3.89)

[a] Results of the classification of compounds obtained from Model 8, $\Delta P\% = [P(\text{active}) - P(\text{inactive})] \times 100$.
[b] Classification.
[c] Observed activity.
[d] Anti-epimastigotes percentage and standard deviation (SS), NC = not classified.

100 μg/ml. Two compounds, **1s** and **4s**, showed appreciable activity at a concentration of 10 μg/ml. Specifically, compound **4s** gave a very interesting result by showing inhibition percentages (%AE) higher than 80% at 100, 10, and 1 μg/ml. Further research will be required to investigate the mechanism of action of these compounds and to evaluate their cytotoxicity at the assayed concentrations. The remaining five compounds which were classified as inactive for the model, showed very low inhibition percentages. For compounds **7s** and **8s**, for instance, 0.0 and 6.9% of inhibition were determined at 100 μg/ml and 0.0 % at other concentrations. According to the model, these are two of the compounds with greater probability of being inactive ones.

This first virtual screening demonstrates the ability of the present *TOMOCOMD* approach to be used for discriminating compounds with potential antitrypanosomal activity from those without this action. These results open, at the same time, a door for the study of several families of heterocyclic compound which appear to be promising sources of antitrypanosomal drugs. Current investigations are being developed in this direction by our research groups.

3. Concluding remarks

The search for effective and rational methodologies for the discovery of new drugs has become a first-line objective in pharmaceutical research. In spite of some criticism, topological indices-based approaches have demonstrated their usefulness in drug discovery processes. *TOMOCOMD* methodology has become an attractive tool to be used in chemical and bioinformatics research. This strategy allowed us to generate a mathematical model with the ability to discriminate antitrypanosomal compounds from inactive ones and to predict, in a rational way, the activity of novel heterocyclic compounds against *T. cruzi*. This family constitutes a starting point for the design and synthesis of more effective and less toxic antitrypanosomal agents.

The current approach can be used in further computational screenings of larger chemical libraries to discover new candidates for antitrypanosomal drugs using a minimum of resources. The interactive and flexible character of the *TOMOCOMD* scheme permits the posterior inclusion of other active and inactive compounds in the training set and the generation, at each step, of more refinished models capable of identifying structural patterns not considered in the present study.

On the basis of the current results we can conclude that the *TOMOCOMD* strategy can be successfully used in the rational search for novel antitrypanosomal compounds.

4. Experimental section

4.1. Computational approach

Calculations were carried out on a PC Pentium-4 2.0 GHz. The *CARDD*-module implemented in the *TOMOCOMD* Software[13] was used in the calculation of total and local non-stochastic quadratic indices. Pauling electronegativities[53] were used as atomic weights (molecular vector's components).

Topological descriptors, molecular walk counts, BCUT descriptors and 2D autocorrelations were calculated by using the DRAGON Software.[47] The molecular structure of each compound was drawn by using the CHEMDRAW software[54] and saved as a mol file. After optimization with the MOPAC software[55] the structures were saved as a hin file and then processed by the DRAGON Software.

4.2. Chemometric method

Linear discriminant analysis (LDA) was performed as implemented in the STATISTICA 5.5 for Windows package.[43] Forward stepwise was fixed as the strategy for variable selection. The quality of the models was determined by examining Wilk's λ parameter (U-statistic), square Mahalanobis distance (D^2), Fisher ratio (F) and the corresponding p-level ($p(F)$) as well as the percentage in training and test sets of global good classification, Matthews' correlation coefficient (MCC), sensitivity, specificity, negative predictive value (sensitivity of the negative category) and false positive rate (false alarm rate). Models with a proportion between the number of cases and variables in the equation lower than 4 were rejected. The statistical robustness and predictive power of the obtained model was assessed using an external prediction (test) set. A leave-group out (10%) cross-validation procedure was also carried out for this propose.

4.3. Parasites and culture procedure

CL strain parasites (clone CL-B5) stably transfected with the *Escherichia coli* β-galactosidase gene (*LacZ*) were used for the assays. Epimastigotes were grown at 28 °C in liver infusion tryptone broth (LIT) with 10% foetal bovine serum (FBS), penicillin and streptomycin.

4.4. Antiepimastigote assay[52]

The screening assay was performed in 96-well microplates with culture that had not reached the stationary phase. Epimastigotes forms, CL strain, were seeded at concentration of 1×10^5 per ml in 200 μl. The plates were then incubated at 28 °C for 72 h with different concentrations of the drugs (100, 10, and 1 μg/ml), at which time 50 μl of CPRG solution was added to give a final concentration of 200 μM. The plates were incubated at 37 °C for 6 h and were then read at 595 nm. Each concentration was assayed three times. In order to avoid drawback, medium, negative and drug controls were used in each test. The anti-epimastigotes percentage (%AE) was calculated as follows: %AE = [(AE-AEB)/(AC-ACB)] × 100, where AE = absorbance of experimental group; AEB = blank of compounds; AC = absorbance of control group; ACB = blank of culture medium. Stock solutions of the compounds to be assayed were prepared in DMSO, with the final concentration in a mixture water/DMSO never exceeding 0.2% of the last solvent.

References and notes

1. World Health Organization (OMS). *Rapport sur la Santé dans le Monde*.WHI: Genève, 1998, 49.
2. Muelas, S.; Suárez, M.; Pérez, R.; Rodríguez, H.; Ochoa, C.; Escario, J. A.; Gómez-Barrio, A. *Mem. Inst. Oswaldo Cruz* **2002**, *97*, 269.
3. Muelas-Serrano, S.; Le-Senne, A.; Fernádez-Portillo, C.; Nogal, J. J.; Ochoa, C.; Gómez-Barrio, A. *Mem. Inst. Oswaldo Cruz* **2002**, *97*, 553.
4. Gonzalbes, R.; Gálvez, J.; Moreno, A.; García-Domenech, R. *J. Pharm. Pharmacol.* **1999**, *52*, 111.
5. Julián-Ortiz, J. V.; Gálvez, J.; Muñoz-Collado, C.; García-Domenech, R.; Gimeno-Cardona, C. *J. Med. Chem.* **1999**, *42*, 3308.
6. McKie, J. H.; Douglas, K. T.; Chan, C.; Roser, S. A.; Yates, R.; Read, M.; Hyde, J. E.; Dascombe, M. J.; Yuthavong, Y.; Sirawaraporn, W. *J. Med. Chem.* **1998**, *41*, 1367.
7. Estrada, E.; Peña, A. *Bioorg. Med. Chem.* **2000**, *8*, 2755.
8. Estrada, E.; Uriarte, E.; Montero, A.; Teijeira, M.; Santana, L.; De Clercq, E. *J. Med. Chem.* **2000**, *43*, 1975.
9. González-Díaz, H.; Marrero-Ponce, Y.; Hernández, I.; Bastida, I.; Tenorio, E.; Nasco, O.; Uriarte, U.; Castañedo, N.; Cabrera, M. A.; Aguila, E.; Marrero, O.; Morales, A.; Pérez, M. *Chem. Res. Toxicol.* **2003**, *16*, 1318.
10. Pérez-González, M.; Dias, L. C.; Morales-Helguera, A.; Morales-Rodríguez, Y.; Gonzaga de Oliveira, L.; Torres-Gómez, L.; González-Díaz, H. *Bioorg. Med. Chem.* **2004**, *12*, 4467.
11. Walters, W. P.; Stahl, M. T.; Murcko, M. *Drug Discovery Today* **1998**, *3*, 160.
12. Drie, J. H. V.; Lajiness, M. S. *Drug Discovery Today* **1998**, *3*, 274.
13. Marrero-Ponce, Y.; Romero, V. *TOMOCOMD* software. Central University of Las Villas. *TOMOCOMD* (*TO*pological *MO*lecular *COM*puter *D*esign) for Windows, version 1.0., 2002.
14. Marrero-Ponce, Y. *Molecules* **2003**, *8*, 687.
15. Marrero-Ponce, Y. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 2010.
16. Marrero-Ponce, Y. *Bioorg. Med. Chem.* **2004**, *12*, 6351.
17. Marrero-Ponce, Y.; Castillo-Garit, J. A.; Torrens, F.; Romero-Zaldivar, V.; Castro, E. *Molecules* **2004**, *9*, 1100.
18. Marrero-Ponce, Y.; Cabrera, M. A.; Romero, V.; Ofori, E.; Montero, L. A. *Int. J. Mol. Sci.* **2003**, *4*, 512.
19. Marrero-Ponce, Y.; Cabrera, M. A.; Romero, V.; González, D. H.; Torrens, F. *J. Pharm. Pharm. Sci.* **2004**, *7*, 186.
20. Marrero-Ponce, Y.; Castillo-Garit, J. A.; Olazabal, E.; Serrano, H. S.; Morales, A.; Castañedo, N.; Ibarra-Velarde, F.; Huesca-Guillen, A.; Jorge, E.; Sánchez, A. M.; Torrens, F.; Castro, E. A. *Bioorg. Med. Chem.* **2005**, *13*, 1005.
21. 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. *J. Comput. Aided Mol. Des.* **2004**, *18*, 615–633.
22. Marrero-Ponce, Y.; Huesca-Guillen, A.; Ibarra-Velarde, F. *J. Theor. Chem. (THEOCHEM)*. **2005**, *717*, 67.
23. Marrero-Ponce, Y.; Montero-Torres, A.; Romero-Zaldivar, C.; Iyarreta-Veitía, I.; Mayón Peréz, M.; García Sánchez, R. *Bioorg. Med. Chem.* **2005**, *13*, 1293.
24. Marrero-Ponce, Y.; Medina-Marrero, R.; Torrens, F.; Martinez, Y.; Romero-Zaldivar, V.; Castro, E. A. *Bioorg. Med. Chem.* **2005**, *13*, 2881.
25. Marrero-Ponce, Y.; González-Díaz, H.; Romero-Zaldivar, V.; Torrens, F.; Castro, E. A. *Bioorg. Med. Chem.* **2004**, *12*, 5331.
26. Milne, G. W. A. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 639.
27. Todeschini, R.; Consonni, V. *Handbook of Molecular Descriptors*; Wiley-VCH: Germany, 2000.
28. Karelson, M. *Molecular Descriptors in QSAR/QSPR*; John Wiley & Sons: New York, 2000.
29. Bonse, S.; Santelli-Rouvier, Ch.; Barbe, J.; Krauth-Siegel, L. *J. Med. Chem.* **1999**, *42*, 5448.
30. De Koning, H. P. *Int. J. Parasitol.* **2001**, *31*, 512.
31. Elhalem, E.; Bailey, B. N.; Docampo, R.; Ujváry, I.; Szajnman, S. H.; Rodíguez, J. B. *J. Med. Chem.* **2002**, *45*, 3984.
32. Avila, J. L.; Rojas, T.; Avila, A.; Polegre, M. A.; Robins, R. K. *Antimicrob. Agents Chemother.* **1987**, *31*, 447.
33. Oliveira, R. B.; Passos, A. P. F.; Alves, R. O.; Romanha, A. J.; Prado, M. A. F.; Dias de Souza Filho, J.; Alves, R. J. *Mem. Inst. Oswaldo Cruz* **2003**, *98*, 141.

34. Carvalho, S. A.; da Silva, E. F.; Santa-Rita, R. M.; de Castro, S. L.; Fraga, C. A. M.. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 5967.
35. Cerecetto, H.; Di Maio, R.; González, M.; Risso, M.; Sagrera, G.; Seoane, G.; Denicola, A.; Peluffo, G.; Quijano, C.; Stoppani, A. O. M.; Paulino, M.; Olea-Azar C.; Basombrio, M. A. *Eur. J. Med. Chem.* **2000**, *35*, 343.
36. Gilbert, I. H. *Biochem. Biophys. Acta* **2002**, *1587*, 249.
37. Du, X.; Guo, Ch.; Hansell, E.; Doyle, P. S.; Caffrey, C. R.; Holler, T. P.; McKerrow, J. H.; Cohen, F. E. *J. Med. Chem.* **2002**, *45*, 2695.
38. Jonckers, T. H. M.; van Miert, S.; Cimanga, K.; Baillo, Ch.; Colson, P.; De Pauw-Gillet, M. C.; van den Heuvel, H.; Caléis, M.; Lemière, F.; Esmans, E. L.; Rozenski, J.; Quirijnen, L.; Maes, L.; Dommisse, R.; Lemière, G. L. F.; Vlietinck, A.; Pieters, L. *J. Med. Chem.* **2002**, *45*, 3497.
39. Negwer, M. *Organic-Chemical Drugs and their Synonyms*; Akademie-Verlag: Berlin, 1987.
40. Mc Farland, J. W.; Gans, D. J. Cluster Significance Analysis. In *Chemometric Methods in Molecular Design*; van Waterbeemd, H., Ed.; VCH publishers: New York, 1995, pp 295–307.
41. Johnson, R. A.; Wichern, D. W. *Applied Multivariate Statistical Analysis*; Prentice-Hall: NJ, 1988.
42. Estrada, E. In *Topological Indices and Related Descriptors in QSAR and QSPR*; Devillers, J., Balaban, A. T., Eds.; Gordon and Breach: Amsterdam, 1999, pp 403–453.
43. STATISTICA, version 5.5; Statsoft Inc., 1999.
44. Baldi, P.; Brunak, S.; Chauvin, Y.; Andersen, C. A.; Nielsen, H. *Bioinform. Rev.* **2000**, *16*, 412.
45. Golbraikh, A.; Tropsha, A. *J. Mol. Graphics Modell.* **2002**, *20*, 269.
46. Wold, S.; Erikson, L. Statistical Validation of QSAR Results. Validation Tools. In *Chemometric Methods in Molecular Design*; van de Waterbeemd, H., Ed.; VCH Publishers: New York, 1995, pp 309–318.
47. Todeschini, R.; Consonni, V.; Pavan, M. DRAGON software version 2.1, 2002.
48. Arán, V. J.; Asensio, J. L.; Ruiz, J. R.; Stud, M. *J. Chem. Soc., Perkin Trans. 1* **1993**, 1119.
49. Ruiz, J. R.; Arán, V. J.; Asensio, J. L.; Flores, M.; Stud, M. *Liebigs Ann. Chem.* **1994**, *1*, 679–684.
50. Arán, V. J.; Flores, M.; Muñoz, P.; Páez, J. A.; Sánchez-Verdú, P.; Stud, M. *Liebigs Ann.* **1996**, 683.
51. Arán, V. J.; Asensio, J. L.; Molina, J.; Muñoz, P.; Ruiz, J. R.; Stud, M. *J. Chem. Soc., Perkin Trans. 1* **1997**, 2229.
52. Vega, C.; Rolón, M.; Martinez-Fernández, A. R.; Escario, J. A.; Gómez-Barrio, A. *Parasitol. Res.* **2005**, *29*, 296.
53. Pauling, L. *The Nature of Chemical Bond*; Cornell University Press: New York, 1939, 2–60.
54. CambrigdeSoft Corpoation. ChemDraw® Ultra. Chemical Structure Drawing standard, 2003.
55. Dewar, M. J. S. *J. Mol. Struct.* **1983**, *41*, 100.