Molecular Diversity (2006) 10: 361–375
DOI: 10.1007/s11030-006-9044-2

Full-Length Paper

Application of linear discriminant analysis in the virtual screening of antichagasic drugs through trypanothione reductase inhibition

Julián J. Prieto, Alan Talevi & Luis E. Bruno-Blanch*
Medicinal Chemistry, Department of Biological Sciences, Exact Sciences Collage, La Plata National University (UNLP), B1900AVV, La Plata, Buenos Aires, Argentina
*(*Author for correspondence, E-mail: lbb@biol.unlp.edu.ar)*

Received 10 January 2006; Accepted 17 May 2006

Key words: Linear Discriminant Analysis, virtual screening, trypanothione reductase inhibitor, Trypanosoma cruzi, Chagas disease

Summary

We have performed virtual screening to identify new lead trypanothione reductase inhibitor (TRI) compounds, enzyme present in *Tripanozoma cruzi*, the agent responsible of Chagas disease. From a training set of 58 compounds, linear discriminant analysis (LDA) was performed using 2D and 3D descriptors as discriminating variables in order to find out which function of descriptors characterizes the active TRI. The values of the statistical parameters F - Snedecor and Wilk's λ for the discriminant function (DF) showed good statistical significance, as long as the rate of success in the prediction for both the training and the test set: 91.38% and 88.63%, in that order. Internal validation through the Leave – Group – Out methodology was performed with good results, assuring the stability of the DF. Afterwards, the DF was applied in virtual screening of 422,367 compounds. The optimum range of values of octanol – water partition coefficient for a compound to develop trypanothione reductase inhibition was applied as a second filtering criteria. 739 structurally heterogeneous drugs of the virtual library were selected as promissory TRI.

Introduction

Motivation

Chagas disease or American Trypanosomiasis, a protozoan infection caused by *Trypanosoma cruzi*, occupies, among protozoan – caused diseases, the third place in the number of deaths per year, after malaria and schistosomiasis [1]. It is a major health problem in Latin America, where current estimates indicate about 20 million people infected with *T. cruzi*, almost 100 million in risk of being infected and 500,000 new cases reported each year. Approximately ten percent of the infected people come from Argentina. This means more than 5 percent of Argentina's population suffers from Chagas disease, which explains our particular concern in the search of new antichagasic agents.

There is neither a vaccine nor any preventive treatment for this parasitosis. Current chemotherapy remains unsatisfactory. Available drugs are benzmidazole and nitrofurans such as nifurtimox. The latter has undergone several rumors of discontinuation [2, 3], probably because of limited markets or the potential risks the pharmaceutical companies may incur because of suspicion of long-term toxicity [4]. Both drugs have significant activity only in the acute and short-term chronic phases. Their efficacy, however, is very low in the established chronic phase, which is prevalent in Latin America and is considered incurable [5]. Their efficacy also varies according to geographical areas, mainly because of differences in drug susceptibility of different T. cruzi strains [2, 6]. Moreover, both medications have important side effects such as anorexia, vomiting, peripheral neuropathy and allergic dermopathy, which can result on treatment discontinuation [2, 5].

Descriptor-based virtual screening arises as one interesting option for researchers from developing countries to discover, in short time and with low costs, promissory drugs in the fight against Chagas.

Virtual screening (vs) in the search of new leads

A drug discovery cycle demands from 10 years or more and the estimate cost in the research and development is $ 1,000 million in research and development cost. The costs and risks of the development of new drugs for the Third World by pharmaceutical industry are out of balance with the perceived limited profits and the long payback period. As mentioned before, even the production of some existing drugs for the treatment of parasitosis is threatened with

discontinuation, because of limited markets and suspicion of long-term adverse effects [4]. The situation is even worse for the academic sector, where funds needed for preclinical and clinical evaluation of drug candidates are not always easy to obtain. Such economic pressure to deliver "best –in-class" drugs has forced drug discovery scientist to develop computational alternatives in the search of new lead drugs, in order to reduce the probability of unsatisfactory results in pre-clinic and clinic evaluation stages [7, 8].

Several methods have been proposed and employed for virtual screening of databases of small organic compounds in order to find tight binders to a given protein target. There are two approaches to this problem. The first of them is "VS by docking" [9], which requires previous knowledge of the 3D structure of the target protein binding site to prioritize compounds by their likelihood to bind to the protein. The other approach involves similarity-based and QSAR VS, where no information on the protein is necessary. Instead, compounds that are known to act by one specific mechanism of action are used as a structural query. The in silico screening procedure extracts compounds from the database according to an appropriate similarity criteria [10]. Similarity-based VS can be carried in two different ways. In the "small molecule alignment" approach, a compound that is known to bind to the target protein, such as the natural substrate or an inhibitor, is used as the reference molecule. During screening, the molecules from the compound database (the test molecules) are superposed onto the reference molecule. The superposition places chemically similar parts of the molecules on top of each other, with a preference on aligning groups able to participate of short range interactions, e.g. H-bonds. Ligand superposition is, however, costly in terms of required computer time. Taking into consideration that chemical databases usually include up to millions of compounds, superimposition is not always applicable to the search of new leads through large compounds databases. This lead us to the "descriptor based screening", which deals with molecular representations that allow for a much more efficient comparison.

For the present work we employed this latter approach, through linear discriminant analysis, in the search for new leaders through more than 422,000 chemical structures. We chose to employ both theoretical 2D and 3D descriptors as independent variables of the DF. There is a considerable debate in literature whether 2D or 3D descriptors are superior, with many articles supporting either opinion [11–13]. Keeping in mind that activity is spatial conformation dependent, 2D descriptors are surprisingly powerful, as revealed by several of the cited bibliographic comparisons. This suggests that in some cases the molecular features critical for activity can be deduced from 2D representations or graphs that do not consider, in a explicit way, conformational concerns. Rather than deciding on one or the other dimensionality approach, we decided on combination of both types of descriptors in order to capture complementary information. This strategy has been suggested as the best approach by many authors [14, 15]. The 2D descriptors used in the present work can be classified as topological indexes (TIs). TIs are molecular descriptors derived from graph-theoretical invariants and accounts for structural information contained in two-dimensional representation of molecules. Although their physicochemical meaning is not always clear, they have some important advantages: they require a low computational cost for their calculation and, as theoretical entities, they can be easily calculated for all the existing, new, and in-development chemical structures without further information than the 2D representation of the molecule [16, 17].

The aim of this study is the development of a new QSAR model based on LDA to predict antitrypanosomal activity through trypanothione reductase (TR) inhibition and, therefore, to employ it in rational selection of new antichagasic drugs. This approach allowed the classification of candidate drugs as active and inactive, identifying 739 promissory candidates among 422,367 compounds previous to correspondent in vitro assays.

Trypanothione reductase: A potential target for antichagasic compounds

TR is an NADPH-dependent flavoenzyme which regenerates a glutathione-spermidine conjugate, N,N-bis(glutathionyl) spermidine, named trypanothione, from its oxidized form (Figure 1). The enzyme is responsible for the maintenance of the redox balance in the Trypanosomatids family, to which the parasites of the genres Trypanosoma and Leishmania belong to. Trypanosomatids do not possess the classic redox system based upon the couple glutathione/glutathione reductase (GR). Despite 41% of homology, TR and GR show almost total discrimination toward their respective substrates. The couple trypanothione/trypanothione reductase has been demonstrated to be essential for the survival of these parasites within the oxidative environment of the host [18]. Moreover, TR is more expressed in infectious than in non–infectious forms of Leishmania [19], which could indicate an association between the enzyme activity and the infective capacity. For these reasons, although many other molecular targets are object of current studies in the fight against Chagas [2], TR is still considered as a promising target for the design of new trypanocidal drugs, as demonstrated by recent papers focalizing in TR and TR inhibitors [18–23]. These studies are based on the hypothesis that it may be possible to find selective medicines which interfere in the trypanothione metabolism of trypanosomatid parasites without distressing the glutathione metabolism of the human host and, therefore, reducing side effects associated to the drug.

Methods

Calculation of molecular descriptors

Dragon computer software was employed to calculate molecular descriptors [24]. Dragon can calculate up to

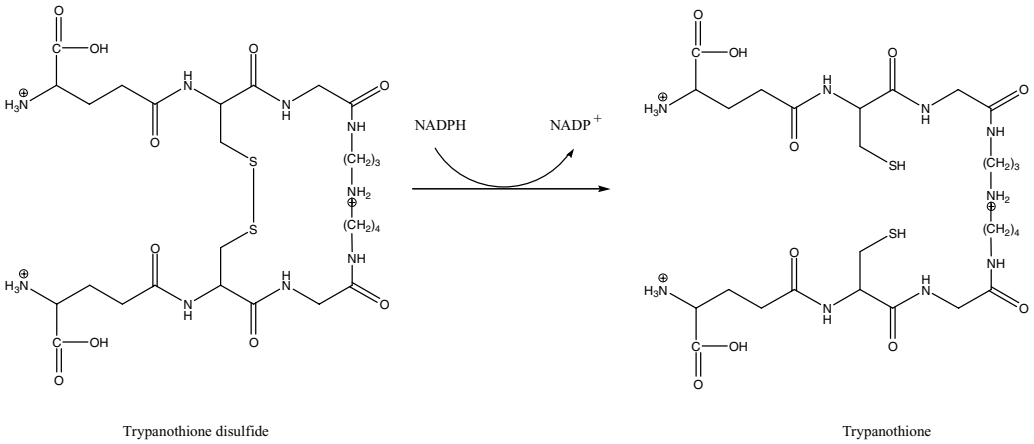

Figure 1. Scheme of the reaction catalyzed by Trypanothione reductase.

1,612 descriptors, among them topological and geometrical descriptors, which are used in this study. Dragon is widely used in QSAR studies for modeling of both physicochemical and biological properties [25–28].

Linear discriminant analysis

The aim of LDA is to find an algorithm capable of distinguishing among two or more categories of objects [29], in this particular case compounds active and inactive against *Trypanosoma cruzi* through TR inhibition. The use of LDA has been applied extensively in drug discovery, mainly by the Las Villas and the Valencia groups [30–34]. Recently, Meneses- Marcel et al. successfully employed this approach in the search of treatment for trichonomiasis, another important parasite- originated disease [35].

The first step in the search of a good linear DF is to use a dataset with good structural diversity. To ensure this point, we selected a dataset of 102 compounds, 50 of them with proven antitrypanosomal activity against extracellular forms of the parasite (epimastigotes and trypomastigotes) and 52 with different or no pharmacological activities. We have considered "active" any compound with IC_{50} against T. cruzi below 100 μmol. The 102 compounds were split, at random, into training and test sets that include 58 and 44 molecules, in that order. The mechanism of action of the selected compounds has been, at least partially, explained through subversive ("turncoat") inhibition of trypanothione reductase, although other simultaneous mechanisms, such as redox cycling with production of hydrogen peroxide and superoxide anion, are admitted in some cases. The training set was composed by 29 active compounds; the remaining 29 (as well as the ones of the test set) compounds have been given other therapeutic uses and belong to the "inactive" category. They were randomly extracted from Merck Index [36]. The 29 compounds classified as "active" included in the training set are showed in Figure 2, while the 29 inactive compounds are presented in Figure 3. The active compounds of the training set were extracted from several bibliographic sources [18, 21, 22, 37–39]. Although most of TR inhibitors are nitroaromatic compounds, we included thirteen compounds without this function, so that the chosen DF was able to select not just nitro compounds as promissory agents. The high proportion of non-nitro drugs among those selected during the subsequent VS process indicates this strategy was successful and that the DF is not biased towards the selection of nitro agents (further discussion of this point can be found in the conclusions). Visual inspection of Figure 2 shows that the active drugs in the training set include at least five analog series, which guarantees structural diversity. Figure 3 reveals that random selection of the inactive compounds of the training set has resulted in structural diversity of the drugs included in the non–active category.

We used BMDP and STATISTICA statistical packages to derive the LDA-based QSAR model [40, 41]. The selection of the descriptor was based on the Fisher-Snedecor F statistical. Stepwise, the discriminating variables (topological indices and 3D descriptors calculated by Dragon) that add the most to the separation of the groups of interest (active and inactive compounds) are entered into (or the variable that adds the least is removed from) the discriminating function (DF). Activity was codified by a dummy variable (class). The variable indicates either the presence (class = 1) or absence (class = −1) of activity against *T. cruzi* through TR inhibition. The DF has the general formula:

$$\text{Class} = a_0x_0 + a_1x_1 + a_2x_2 + \ldots + a_nx_n + c \qquad (1)$$

where $x_0, x_1, \ldots, x_n$ represent the chosen descriptors and $a_0, a_1, \ldots, a_n$ represent the coefficients of the classification function, determined by the least-squares method. We employed the principle of parsimony, choosing a DF with high statistical significance but having as few parameters as possible. The proportion between the number of cases and the number of variables included in the equation was observed in order to avoid over-fitting and chance correlations. The analyzed statistical parameters were Wilk's λ statistic, Fisher ratio (F) and the p-level (p). Internal validation through

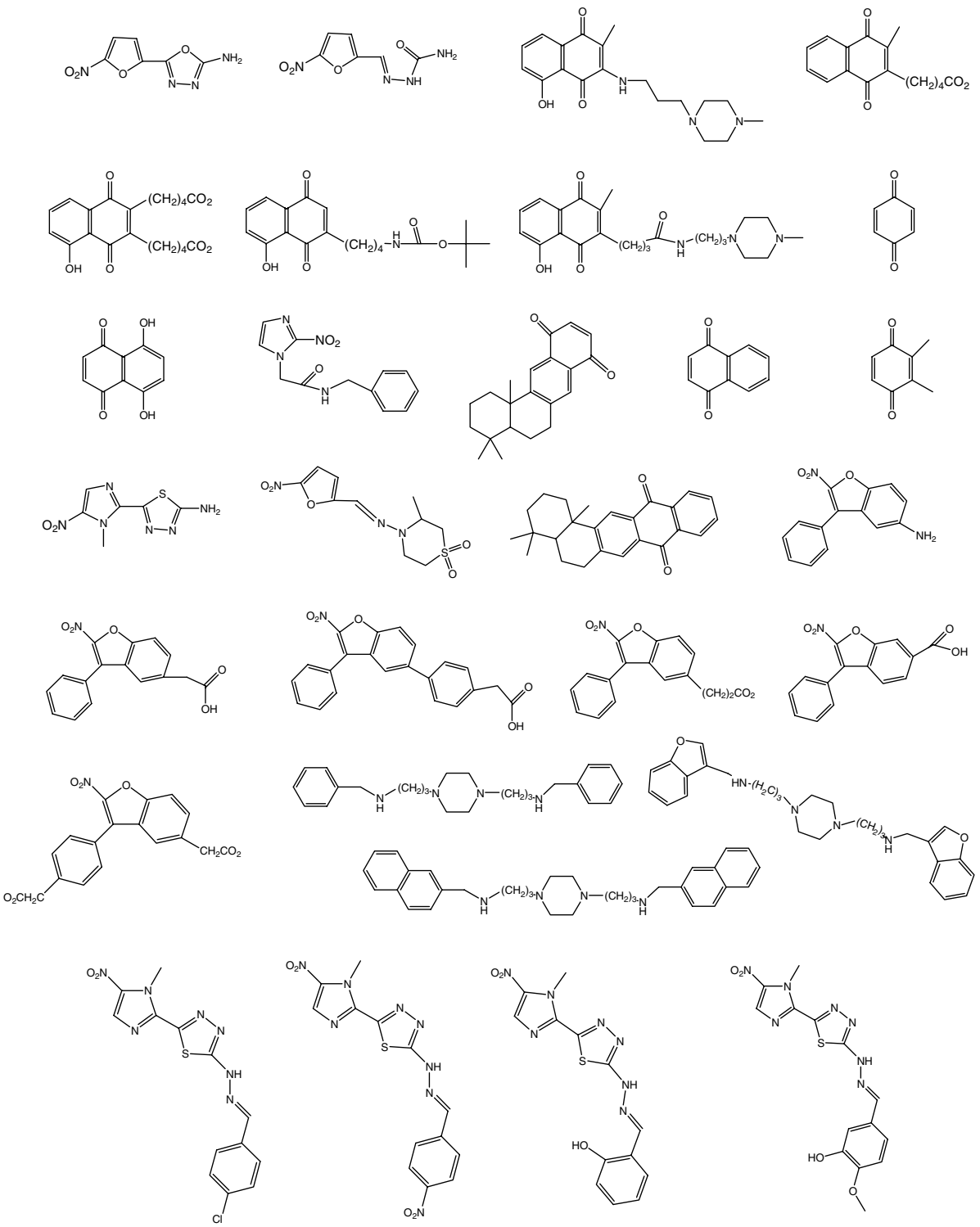

Figure 2. Structures of the 29 active compounds included in the training set.

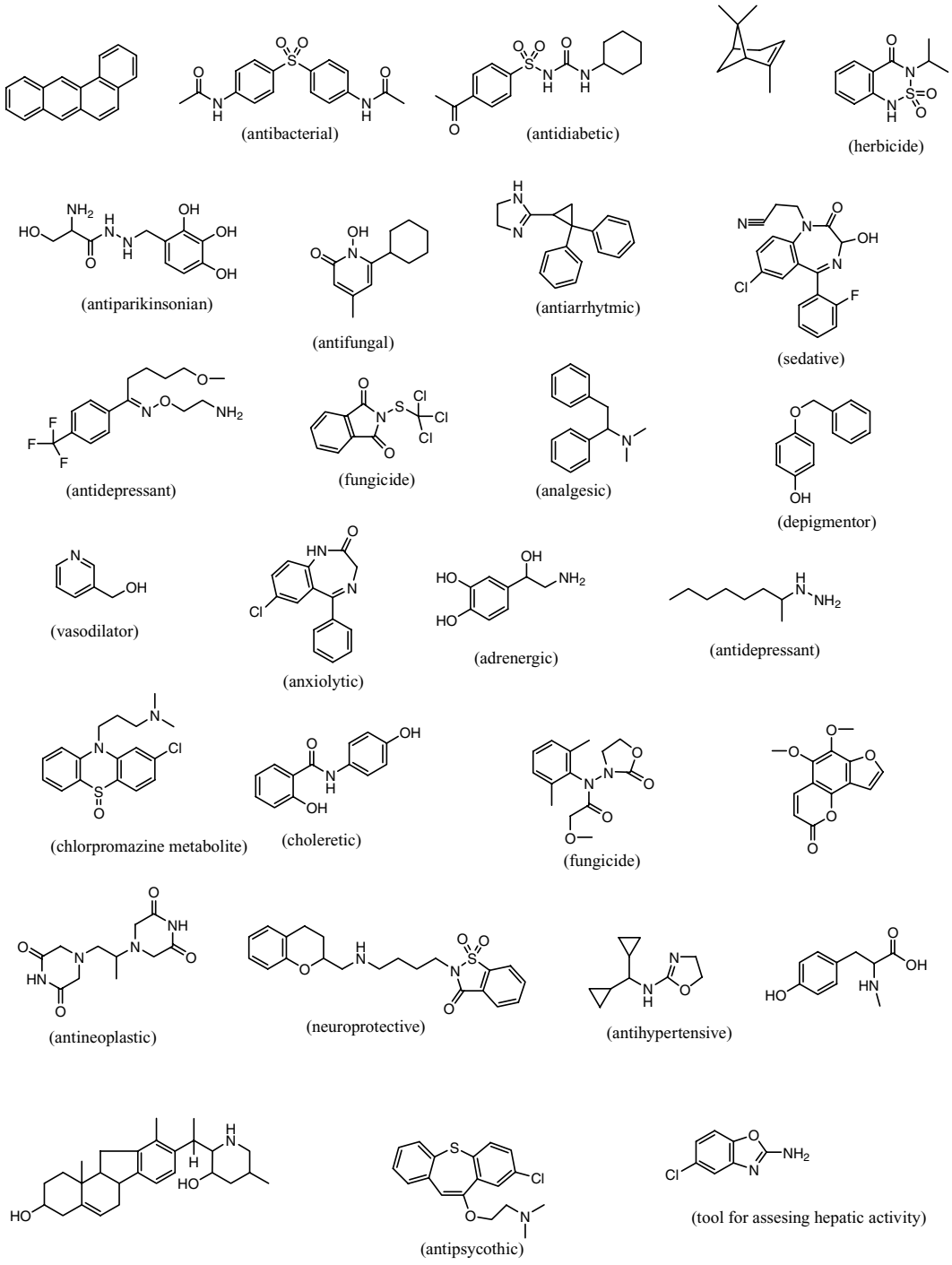

Figure 3. Structures of the 29 inactive compounds included in the training set. Structural diversity can be appreciated. In parenthesis, below each compound, we present its therapeutic use according to Merck Index 13th.

Leave-Group-Out methodology was also performed, in order to assure no molecule of the training set has excessive influence over the generated DF. We inspected the percentage of good classifications in both groups (active and inactive molecules) as well as the total percentage of good classifications. Model predictability was tested with an external prediction series of compounds (those referred as the test set) which were not used to develop the DF. A posterior probability P (%) was assigned to each compound to score its biological activity.

Pharmacological distribution diagrams (PDD)

The generated DF is capable of describing pharmacological activity patterns, but also non-activity patterns. A PDD is a frequency distribution diagram of a dependent variable (in our case, value of the DF for each compound of the training and the test sets) in which the ordinate represents the expectancies of this variable for every interval. The expectancies of the DF are defined as the probability that a compound will be active or inactive for a range of values of the DF. They are obtained by means of the expressions indicated below [42]. The 100 that appears in the discriminator prevents from dividing by zero.

$$Ea = \frac{\text{percentage} - \text{of} - \text{actives}}{(\text{percentage} - \text{of} - \text{inactives} + 100)} \quad (2)$$

$$Ei = \frac{\text{percentage} - \text{of} - \text{inactives}}{(\text{percentage} - \text{of} - \text{actives} + 100)} \quad (3)$$

The main advantage of these diagrams is that they allow the visual determination of the intervals of the DF where there is a maximum probability of finding new active compounds and a minimum of encountering inactive ones. Therefore, they facilitate the election of the cutoff value that one will apply in the virtual screening procedure to differentiate active from inactive compounds.

Application of the df in virtual screening and log p value as a second filtering criteria

The validated QSAR model was then applied to the prediction of activity of more than 422,000 compounds. The descriptors included in the model, as well as the Moriguchi log P (mlog P) [43], were calculated in Dragon and it was verified that the descriptors values fall within the chemical space defined by the training set. The Moriguchi log P was then applied as a secondary filtering criteria.

In order to be effective in the chronic stage of Chagas disease, it is essential that the antichagasic drug reach the intracellular form of the parasite. This is a matter of great importance, since that is the phase of Chagas that currently lacks of adequate treatment. When it comes to an oral treatment, an antichagasic agent acting through TR inhibition must be able to penetrate the intestinal membrane of the human host, the cellular membrane of the affected organs and the parasite membrane before producing its therapeutic effect. Therefore, octanol–water partition coefficient (log P) is a critical parameter regarding antichagasic activity. Daunes and D'Silva have pointed optimum log P value to develop antichagasic activity through TR inhibition in 5.8, with a log P range from 4 to 7 required for good parasite membrane penetration [44, 45] while Aguirre et al. [46] fixed this value in about 2.3 in a rational design study for trypanothione reductase inhibitors.

In the other hand, the Lipinski rule of five states that poor absorption or permeation of a drug is more likely when the molecular weight of the compound is over 500, there are more than five hydrogen bond donors (expressed as the sum of hydroxyl and amino groups present in a molecule), the calculated octanol–water partition coefficient (clog P) is over 5 or the mlog P is over 4.15, and there are more than 10 hydrogen bond acceptors. Compound classes that are substrates for biological transporters are exceptions to Lipinski's rule. If two parameters are out of the range then a poor absorption or permeability is likely to occur [47, 48]. Besides, lipophilicity is a possible contributory factor to the toxicity of a drug. The relationship between log P and toxicity is further complicated by the increased residence time in lipid-containing structures of compounds with high log P. Compounds with very high log P can accumulate within the organism and cause long-term adverse effects [49].

However, a recent study shows that from 1,791 synthetic oral drugs approved or marketed from 1937 to the present, 152 (about 8.5 percent) have clog P values above 5 (which is considered equivalent to a mlog P value of 4.15) [50]. Furthermore, Ghose et al. [48] have noticed that mean log P values for some specific types of drugs can be below or above the preferred log P range (in general therapeutic agents) of 1.3 to 4.1. Many of the exceptional drugs with clog P values above 7.0 signaled by Ghose et al. are antiparasitary agents (among them: dymanthine, zilantel, bisbendazole, halofantrine, lapinone, menoctone, aminoquinol and chloramphenicole palmitate). Consequently, there is now considerable interest in the potential of lipid formulations for oral administration [51]. Lipid systems can assure a superior bioavailability of drugs with high log P, with two prominent consequences: they allow pharmacological activity to be achieved with lower doses and they allow utilization of lipophilic drugs whose high potency allows the use of very low doses, compensating the higher toxicity linked to high log P values. In other words, there are currently many pharmacological and pharmaceutical reasons not to straightforward discard lipophilic drugs that have shown high potency.

From the exposed considerations it is clear that choosing and adequate range of log P values to use as secondary filtering criteria for antichagasic drugs is not a simple matter. On this basis we decided to select, after VS, only those compounds with Moriguchi log P (mlog P) value between 1.5 and 5.0, allowing some slight flexibility in the Lipinski rule regarding log P but carefully watching that none of

selected compounds infringes the other parameters in the "rule of five".

Results

Discriminant function

Several DFs were tested. Among those with lowest Wilk's λ value, we selected the one with the best performance in the external validation, that is, the DF with the highest percentage of good classifications in the test set. The chosen function was:

$$\begin{aligned} \text{Class} = {} & -4.8095 - 42.7723\,\text{JGI6} - 0.0376\,\text{QXXp} \\ & + 2.6058\,\text{PJI3} + 24.2097\,\text{JGI4} + 0.0106\,\text{ZM2V}. \\ \text{N} = 58 \quad & \text{F} = 17.026 \quad \text{U} - \text{statistic (Wilk's } \lambda) \\ & = 0.379 \quad p < 0.0000 \end{aligned}$$

This equation is statistically significant above 99.99% level. Concerning the variables all are significant above 99%. The Wilks' statistic for the overall discrimination can take values in the range of 0 (perfect discrimination) to 1 (no discrimination). The value obtained for this statistic (0.379) shows the selected DF is able to discriminate among active and inactive compounds against Chagas through TR inhibition. This can be visually corroborated by inspection of Figure 4. The set of descriptors showed low values of the mean correlation coefficient of the intercorrelation matrix, $R_{IM} = 0.353$. The weakest and strongest correlation coefficients being $R_w = 0.219$ and $R_S = 0.626$. This means all variables included in the function are weakly correlated (far below the maximum linear correlation coefficient of $R = 0.9$ admitted by the non–redundant descriptors principle [25, 52]). JGI4 and JGI6 are Galvez Mean Topological Charge Indexes of fourth and sixth orders; [53] QXXp represents the Qxx COMMA2 value weighted by atomic polarizability [54]; PJI3 denotes the 3D Petitjean Shape Index [55] and ZM2V stands for the Second Zagreb index by Valence Vertex Degrees [56]. QXXp and PJI3 are geometrical descriptors, while JGI4, JGI6 and ZM2V are topological in nature.

The results of compound classification are showed in Table 1. The model correctly classifies 91.38 percent of the compounds from the training set and 88.63 percent of the test set (external validation). The rate of "false actives" is similar to 10 percent for the training set. It is desirable for the number of false active compounds to be as low as possible because this number represents inactive compounds that would be sent to biological assays with the consequent loss of time and resources. The obtained value implies about one of ten compounds send to biological assays will show no activity against *T. cruzi*, which would be an acceptable rate of failure. Nevertheless, as explained below, the use of PDDs to

Table 1. Classification of compounds from the training and test sets. When generic name does not exist, the chemical name is used. Posterior probabilities and predicted class associated to each compound are showed. Only five compounds (out of 58) are misclassified in the training set. Compounds with ΔP = (posterior probability of being active – posterior probability of being inactive) < 5% are considered unclassified (U).

Compound	Prob%	Class
Training set active compounds		
(+)-(4aS,12bS)-4,4,12b-Trimethyl-1,2,3,4,4Â ,5,6 12b-octahydro- benzo[a]anthracene-8,11-dione	50.66	U
p-Naphthoquinone	82.02	+
Naphthazarin	96.69	+
p-Benzoquinone	97.98	+
2,3-dimethyl-p-benzoquinone	68.22	+
(+)-(4aS,14bS)-4,4,14b-Trimethyl-1,2,3,4,4a,5,6 14b-octahydro-benzo[a]naphthacene-8,13-dione	89.22	+
2-(5-amino-1,3,4-oxadiazolyl)5-nitrofurane	97.38	+
2-methyl-3-(N-boc-4-aminobutyl)juglone	99.23	+
3(N-boc-5-aminopentyl)menadione	96.31	+
3-[(N-1-propyl-3-methyl-piperazide) propanamideplumbagine	96.21	+
3-(1-propyl-3-methylpiperazinyl)menadione	97.42	+
3-(4-carboxybutyl)menadione	89.62	+
2-formyl-5-nitrofurane semicarbazine	70.41	+
Benzmidazol	94.27	+
Megazol	47.17	–
Nifurtimox	92.19	+
5-amino-3-phenyl nitrobenzofuran	54.52	+
6-carboxy-3-phenyl nitrobenzofuran	99.46	+
1,4-Bis{3-[N-(benzyl)amino]propyl}piperazine	8.98	–
1,4-Bis{3-[N-(napht-2-ylmethyl)amino] propyl} piperazine	94.21	+
1,4-Bis{3-[N-(benzofur-3-ylmethyl)amino] propyl}piperazine	95.52	+
3-(4-carboxymethylphenyl) nitrobenzofuran	98.02	+
5-carboxyethyl-3-phenyl nitrobenzofuran	99.53	+
5-carboxymethyl-3-phenyl nitrobenzofuran	94.94	+
5-(4-carboxymethyl)phenyl-3-phenyl nitrobenzofuran	99.80	+
1-methyl-2-[5-(4-chlorophenylhidrazone)-1,3,4-thiadiazolyl]-5-nitroimidazol	99.53	+
1-methyl-2-[5-(4-nitrophenylhidrazone)-1,3,4-thiadiazolyl]-5-nitroimidazol	99.92	+
1-methyl-2-[5-(2-hydroxyphenylhidrazone)-1,3,4-thiadiazolyl]-5-nitroimidazol	99.78	+
1-methyl-2-[5-(3-hydroxy-4-metoxi-phenyl hidrazone)-1,3,4-thiadiazolyl]-5-nitroimidazol	99.94	+
Training set inactive compounds		
1,2-Benzanthracene	41.09	–
Acedapsone	6.04	–
Acetohexamide	0.35	–
Alpha-pinene	45.09	–
Benserazide	39.95	–
Bentazon	0.29	–
Ciclopirox	1.02	–
Cifenline	6.44	–
Cinolazepam	31.87	–
Fluvoxamine	7.63	–
Folpet	0.80	–
Lefetamine	0.21	–
Molindone	2.61	–
Monobenzone	25.66	–
Nicotinyl alcohol	3.11	–

(*Continued on next page*)

Table 1. (*Continued*)

Compound	Prob%	Class
Nordazepam	0.33	–
Norepinephrine	11.53	–
Octamoxin	0.00	–
Opromazine	0.27	–
Osalmid	69.03	+
Oxadixyl	11.94	–
Pimpinellin	8.29	–
Razoxane	36.93	–
Rilmenidine	78.57	+
Surinamine	0.83	–
Repinotan	0.47	–
Veratramine	1.40	–
Zotepine	4.22	–
Zoxazolamine	65.49	+
Test set active compounds		
3-(4-carboxybutyl)plumbagin	92.97	+
3-(1-propyl-3-(3-propylamino)piperazinyl) menadione	67.21	+
3(N-boc-4-aminobutyl)menadione	88.36	+
3(N-boc-3-aminopropyl)menadione	84.66	+
3-(5-carboxypentyl)plumbagin	92.69	+
5-[4-(2-hydroxiehtyl)phenyl]-3-phenylnitrobenzofuran	88.58	+
Juglone	94.77	+
3-(5-carboxypentyl)menadione	84.27	+
5-(2-hydroxiethyl)-3-phenylnitrobenzofuran	74.47	+
(+)-(4aS,12bS)-4,4,12b-Trimethyl-1,2,3,4,4a,5,6,12b-octahydro-benzo[a]anthracene-8,11-dione	40.70	–
(-)-(7R,7aS,11aS)-7-[1,3]Dioxolan-2-yl-7a,11,11-trimethyl-7,7a,8,9,10,11,11a,12 octahydro-benzo[a]anthracene-1,4-dione	3.66	–
6-carboxymethyl-3-phenylnitrobenzofuran	97.44	+
5-carboxy-3-phenylnitrobenzofuran	94.56	+
5-carboxymethyl-3-(3-bromophenyl) nitrobenzofuran	96.08	+
5-(2-carboxyehtyl)-3-phenylnitrobenzofuran	95.08	+
5-carboxymethyl-3-(2-fluorophenyl)nitrobenzofuran	98.84	+
1-methyl-2-[5-(4-bromophenylhidrazone)-1,3,4-thiadiazolyl]-5-nitroimidazol	97.18	+
1-methyl-2-[5-(3,4-dihydroxyphenylhidrazone)-1,3,4-thiadiazolyl]-5-nitroimidazol	99.43	+
1-methyl-2-[5-(4-hydroxyphenylhidrazone)-1,3,4-thiadiazolyl]-5-nitroimidazol	99.82	+
2-phenyl-5-(3,4-dihydroxyphenylhidrazone)-1,3,4-thiadiazole	87.19	+
(+)-(4aS,12bS)-4,4,9,10,12b-Pentamethyl-1,2,3,4,4a,5,- 6,12b-octahydro-9-hydroxi- benzo [a]anthracene-8,11-dione	92.74	+
Test set inactive compounds		
Clenbuterol	22.88	–
Clobenfurol	43.37	–
Dexamethasone	60.91	+
Dexanabinol	0.21	–
Dexetimide	2.36	–
Dexpanthenol	0.63	–
Gossypol	88.04	+
Grepafloxacin	39.00	–
Guaiacol	36.36	–
Guanadrel	4.83	–
Halazepam	75.64	+
Sulbenox	1.14	–
Sulbentine	8.50	–

(*Continued on next page*)

Table 1. (*Continued*)

Compound	Prob%	Class
Tiamenidine	1.77	–
Tiaprost	3.54	–
Tiazofurin	30.35	–
Trimethadione	1.40	–
Trithiozine	5.41	–
Vedaprofen	1.04	–
Venlafaxine	0.06	–
Xenbucin	0.51	–
Zaltoprofen	39.89	–
2-phenyl-5-(3,5-di-t-buthyl-4-hydroxyphenylhidrazone)-1,3,4-thiadiazole	6.62	–

define the cutoff value to be used in the VS allows to reduce even more the "false active" rate. The correct classification of 2-phenyl-5-(3,5-di-t-buthyl-4-hydroxyphenylhidrazone)-1,3,4-thiadiazole represents further evidence of the good classification capability of the DF, since this compound is actually inactive against *T. cruzi* but structurally similar to other 1,3,4-thiadiazole-2-arylhydrazone active derivatives that are included (and correctly classified) in both the training and test sets.

Robustness of the selected DF was assessed through cross validation through the leave-group-out methodology. The validation of an internal set through this methodology is illustrated in Table 2. Six runs were carried out. In each run, ten compounds (both from the active and inactive category) were extracted from the training set to the test set, generating one new DF with the remaining 48 compounds of the training set. Wilks' λ values for each equation are displayed, as well as the percentage of total good classifications for both the new training and the test sets in each run. The average percentage of success for the training set is 89.93% and for the test set 88.55%. The results were similar to the ones obtained with the actual DF (91.38 and 88.63%), which indicates the

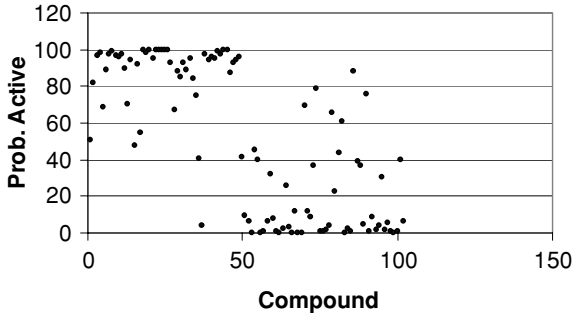

Figure 4. Plot of the Probability of belonging to the active group for each compound in both the training and the test sets. Active compounds are numbered below 50. Visual inspection of the plot reveals that the DF is able to separate the compounds in two well differenced zones: about 90 percent of the active compounds present probabilities of being active above 90%, while most of the non–active compounds (83%) showed posterior probability values below 40 percent.

Table 2. Results of internal validation for DF. It can be appreciated that neither the statisticals λ and F nor the percentages of good classifications vary significantly with the removal of ten inactive and active compounds, in each leave-group-out run, from the training set.

Number of Leave one Out Cross validation	N	Wilks 'λ	F	p	Percentage of total correct classifications in training set	Percentage of good classifications in test set	Compounds removed from the original training set
1	48	0.364	14.70	<0.0000	89.58	88.89	Lefetamine; molindone; monobenzone; nicotinyl alcohol; nordazepam; 2-(5-amino-1,3,4-oxadiazolyl) 5-nitrofurane; megazol; nifurtimox; 1-methyl-2-[5-(4-chlorophenylhidrazone)-1,3,4-thiadiazolyl]-5-nitroimidazol; 1,4-Bis{3-[N-(benzyl)amino]propyl} piperazine
2	48	0.365	14.60	<0.0000	89.58	88.89	Bentazon; ciclopirox; cifenline; cinolazepam; fluvoxamine; 2-formyl-5-nitrofurane semicarbazine; benzmidazol; (+)-(4aS,14bS)-4,4,14b-trimethyl-1,2,3,4,4a,5,6 14b-octahydro-benzo[a]naphthacene-8,13-dione; p-Naphthoquinone; 3-(4-carboxybutyl)menadione
3	48	0.406	12.29	<0.0000	89.58	90.74	Rilmenide; surinamide; veratrapine; zotepine; 14b-octahydro-benzo[a]naphthacene-8,13-dione; benzmidazole; 2-formyl-5-nitrofurane semicarbazine; (+)-(4aS,12bS)-4,4,12b-Trimethyl-1,2,3,4,4Â ,5,6 12b-octahydro- benzo[a]anthracene-8,11-dione; 1-methyl-2-[5-(2-hydroxyphenylhidrazone)-1,3,4-thiadiazolyl]-5-nitroimidazol; 6-carboxy-3-phenyl nitrobenzofuran
4	48	0.356	14.82	<0.0000	89.58	88.89	pimpinelline, octamoxin, opromazine, osalmid, oxadixyl, norepinephrine, 5-carboxyethyl-3-phenyl nitrobenzofuran, 3-(4-carboxybutyl)menadione, 2-formyl-5-nitrofurane semicarbazine; 1-methyl-2-[5-(3-hydroxy-4-metoxi-phenyl hidrazone)-1,3,4-thiadiazolyl]-5-nitroimidazol
5	48	0.358	15.09	<0.0000	91.66	87.04	Alpha-pinene; ciclopirox; folpet; razoxane; repinotan; megazol; 1-methyl-2-[5-(4-chlorophenylhidrazone)-1,3,4-thiadiazolyl]-5-nitroimidazol; p-Benzoquinone; 6-carboxy-3-phenyl nitrobenzofuran; 2-methyl-3-(N-boc-4-aminobutyl)juglone
6	48	0.385	13.43	<0.0000	89.58	87.04	1,2-Benzanthracene; acedapsone; acetohexamide; zotepine; zoxazolamine; 1-methyl-2-[5-(4-nitrophenylhidrazone)-1,3,4-thiadiazolyl]-5-nitroimidazol; 5-carboxyethyl -3-phenyl nitrobenzofuran; 1-methyl-2-[5-(2-hydroxyphenylhidrazone)-1,3,4-thiadiazolyl]-5-nitroimidazol; 3-(4-carboxybutyl)menadione

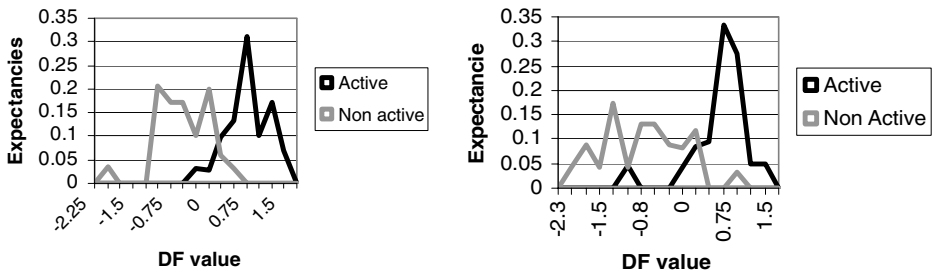

Figure 5. PDD for training set (left) and test set (right The good separation of the groups according to the DF values is clear. The cutoff value to apply in the VS was derived from the PDD of both sets. Compounds with DF function between 0.5 and 2.0 were considered potentially active.

Table 3. We present the molecule number (corresponding to the number given to the molecules in Figure 6a and b), molecules names, DF and mlog P values of the 40 structures selected through the VS process with highest DF values (over the interval 0.5–2.0). Generic names are presented when available; otherwise, chemical names are showed.

Molecule Number	Molecule name	DF	mlog P
1	3-(4-chlorophenyl)-5-[1-(4-nitro-phenyl)-5-(trifluoromethyl)pyrazol-4-yl]-1,2,4-oxadiazole	1.99	3.38
2	2-(trifluoromethyl)-5-[[2-(trifluoromethyl)-3H-benzoimidazol-5-yl]oxy]-3H-benzoimidazole	1.95	2.64
3	6-methoxy-2-(1,1,2,2,3,3,4,4-octafluorobutyl)chromen-4-one	1.94	3.16
4	2-[[4-amino-5-(2-fluorophenyl)-1,2,4-triazol-3-yl]sulfanyl]-N-(2,3,4,5,6-pentafluorophenyl)acetamide	1.93	2.56
5	2,2,2-trifluoro-N-[2-[4-[(2,2,2-trifluoroacetyl)amino]phenyl]-3H-benzoimidazol-5-yl]acetamide	1.92	3.16
6	N-[4-(1,3-dihydrobenzoimidazol-2-yl)phenyl]-2,3,4,5,6-pentafluoro-benzamide	1.84	4.35
7	3-[4-(trifluoromethoxy)phenyl]-1-[[4-(trifluoromethoxy)phenyl] carbamoylamino]urea	1.83	3.04
8	1,2,4,5-tetrafluoro-3-(4-nitrophenoxy)-6-(trifluoromethyl)benzene	1.82	4.59
9	N-[4-(1,3-dihydrobenzoimidazol-2-yl)phenyl]-2,3,5,6-tetrafluoro-4-methoxy-benzamide	1.81	3.67
10	[4-[[3-(5-chlorobenzooxazol-2-yl)phenyl]iminomethyl]phenyl] furan-2-carboxylate	1.81	3.34
11	3-[3,5-bis(trifluoromethyl)phenyl]-5-(oxazol-5-yl)-1,2,4-oxadiazole	1.74	3.36
12	(4E)-2-(4-nitrophenyl)-4-[[5-[3-(trifluoromethyl)phenyl]-2-furyl]methylidene]-1,3-oxazol-5-one	1.73	3.82
13	1-[(2,3,5,6-tetrafluoropyridin-4-yl) amino]-3-[4-(trifluoromethyl)phenyl]urea	1.70	2.71
14	[2-nitro-4-(trifluoromethyl)phenyl] 3-(trifluoromethyl)benzoate	1.69	4.89
15	ethyl-2,3,5,6-tetrafluoro-4-(2,3,4,5,6-pentafluorophenoxy)benzoate	1.67	4.81
16	3-(2,6-Dinitro-4-trifluoromethyl-phenoxy)-benzo[c]chromen-6-one	1.66	4.08
17	3-(4-methyl-1,3-thiazol-2-yl)-7-[(2,3,4,5,6-pentafluorophenyl) methoxy]chromen-2-one	1.62	4.05
18	[3-(4-chlorophenyl)-4-oxo-2-(trifluoromethyl)chromen-7-yl] furan-2-carboxylate	1.55	3.40
19	[3-(4-chlorophenyl)-4-oxo-chromen-7-yl] 4-methyl-3,5-dinitro-benzoate	1.51	3.75
20	2,3,4,5,6-pentafluorophenyl) carba-moylmethyl 2,4-dihydroxybenzoate	1.50	2.68
21	6-nitro-2-(trifluoromethyl)chromen-4-one	1.49	1.91
22	[4-(4-nitrobenzoyl)oxyphenyl] 4-nitrobenzoate	1.49	4.15
23	2,3,5,6-tetrafluoro-4-[4-(2,3,5,6-tetrafluoropyridin-4-yl)oxyphenoxy]pyridine	1.48	2.61
24	2-[[4-amino-5-(4-methoxyphenyl)-1,2,4-triazol-3-yl]sulfanyl]-N-(2,3,4,5,6-pentafluorophenyl)acetamide	1.47	1.91
25	(4E)-2-(2-chloro-4-nitro-phenyl)-4-[[5-(3-nitrophenyl)-2-furyl]methylidene]-1,3-oxazol-5-one	1.45	3.58
26	7-(2,3,4,5-tetrafluorophenyl)-2-(trifluoromethyl)-6-thia-1,3,4,8-tetrazabicyclo[3.3.0]octa-2,4,7-triene	1.45	2.31
27	[2-(4-methoxyphenyl)-4-oxo-chromen-6 -yl] 4-chloro-3-nitro-benzoate	1.44	3.19
28	[(2E)-2-(benzo[1,3]dioxol-5-ylmethylidene)-3-oxo-benzofuran-6-yl] benzo[1,3]dioxole-5-	1.39	1.97
29	[2-nitro-4-(trifluoromethyl)phenyl] 2,6-difluorobenzoate	1.39	4.54
30	[3-(2-chlorophenyl)-4-oxo-chromen-7-yl] 4-methyl-3,5-dinitro-benzoate	1.38	3.74
31	1-[[5-(trifluoromethoxy)1H-indole-2-carbonyl]amino]-3-[4-(trifluoromethoxy)phenyl]urea	1.37	2.25
32	1-[(2,3,5,6-tetrafluoropyridin-4-yl)amino]-3-[2-(trifluoromethyl)phenyl]urea	1.37	2.71
33	2,2,2-trifluoro-N-[2-[4-[5-[(2,2,2-trifluoroacetyl)amino]-3H-benzoimidazol-2-yl]phenyl]-3H-benzoimidazol-5-yl]acetamide	1.35	3.26
34	1-bromo-2,3,5,6-tetrafluoro-4-(2,3,5,6-tetrafluoro-4-methoxy-phenoxy)benzene	1.34	4.69
35	octafluorobenzidine	1.34	4.09
36	[1-amino-2,2-difluoro-2-[(2S)-2,3,3,4,4,5,5-heptafluorooxolan-2-yl]ethylidene]-benzothiazol-2-yl-azanium	1.31	2.78
37	N1', N2'-bis(3-nitrobenzoyl)ethanedihydrazide	1.30	1.91
38	N-[2-[3-naphthalen-2-yl-5-(trifluoro-methyl)pyrazol-1-yl]-2-oxo-ethyl]-5-nitro-3H-imidazole-4-carboxamide	1.29	1.97
39	N-[4-(5-chlorobenzooxazol-2-yl)phenyl]-1-(6-nitrobenzo[1,3]dioxol-5-yl)methanimine	1.27	2.83
40	N-[[4-(carbamoylmethoxy)phenyl] methylideneamino]-4-(1,1,2,2-tetrafluoroethoxy)benzamide	1.26	2.39

(*Continued on next page*)

(*Continued*)

stability of the model. Table 2 also shows which compounds were removed for each leave– n–out run.

Virtual screening

After visual inspection of PDDs for training and test sets (Figure 5), the range 0.5–2.0 for the DF values was employed as criteria to decide if a compound was promissory as an anticha-

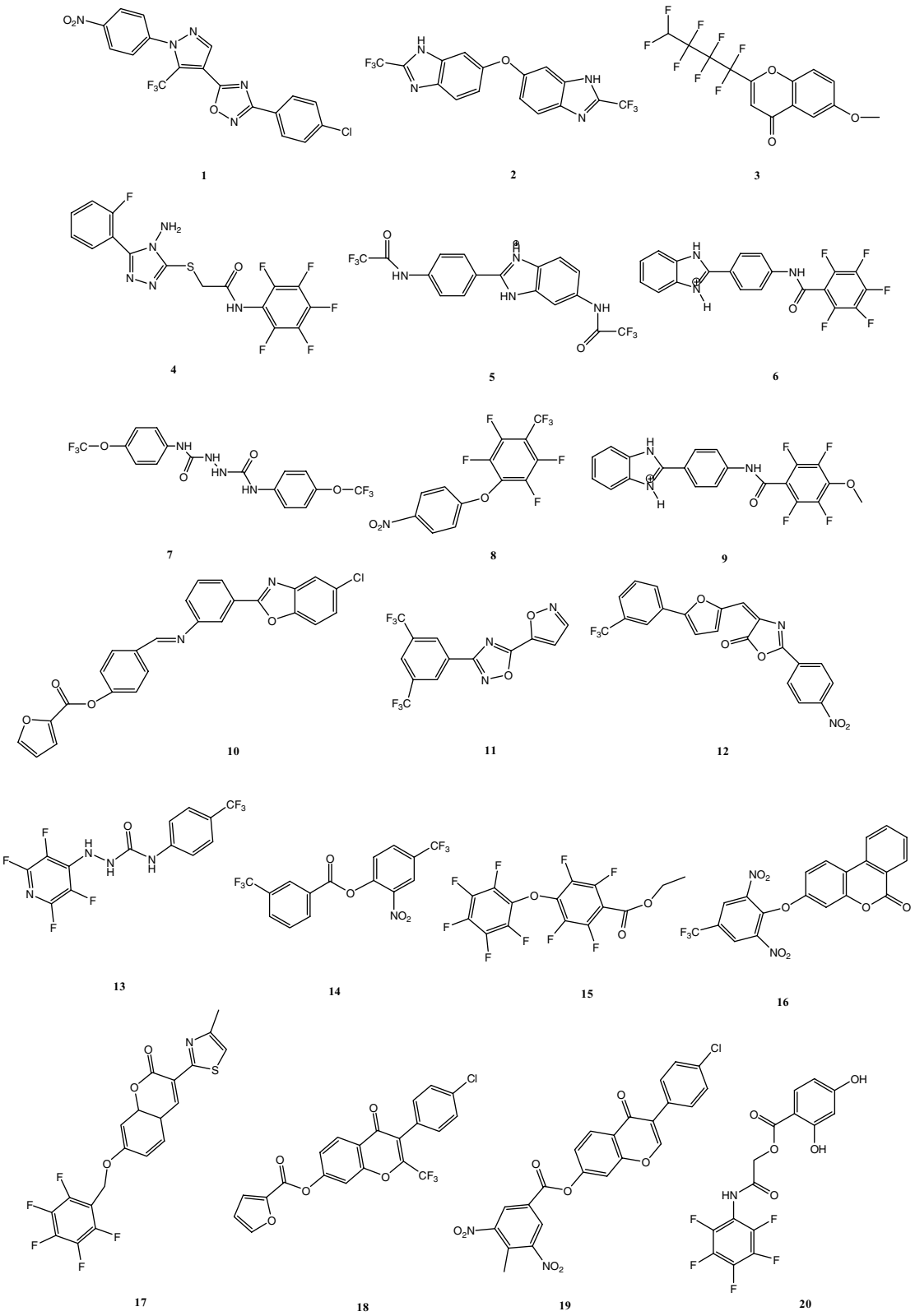

Figure 6. (a) Structures 1–20 of the 40 drugs selected through VS with highest DF values in the interval 0.5–2.0. The numeration is in correspondence with the numbers of each drug in Table 3.

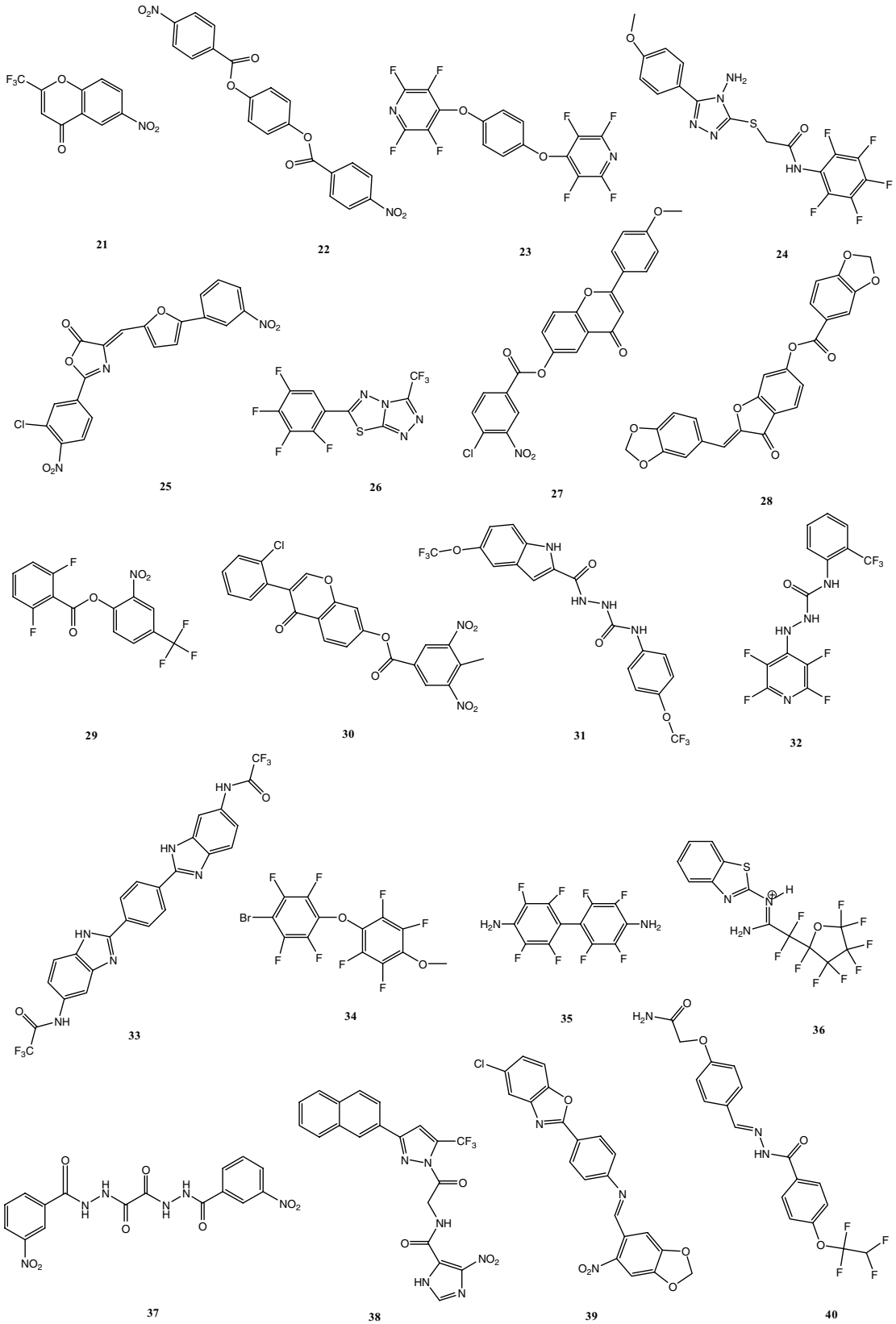

Figure 6. (b) Structures 20–40 of the 40 drugs selected through VS with highest DF values in the interval 0.5–2.0. The numeration is in correspondence with the numbers of each drug in Table 3.

gasic agent (or not). Although there are high probabilities for a compound with DF value above 0 to be active, we preferred to minimize the "false active" rate, choosing an interval of DF values with low overlapping in respect to the non-active plot. This way we try to assure the bioassayed compounds are active. As already mentioned, mlog P value was employed as a second filtering criteria, giving priority to those compounds with mlog P between 1.5 and 5.0.

739 compounds were selected as promissory antichagasic agents through TR inhibition. The DF and mlog P values of the forty compounds with higher log P value are presented in Table 3. Their structures are shown in Figure 6a and b.

Conclusions

From the results that have been exposed, we can formulate several observations:

- The internal validation of the DF showed good results. It is remarkable that every one of the models generated in each of the six leave-group-out runs presented similar results to those of the original model obtained with the complete 58-molecules training set. This seems to indicate the DF presents good stability.
- The external validation showed good results, with nearly 90% good classifications in the test sets. This indicates only one in ten of the molecules selected in the VS would be a "false active" if we consider that molecules with a DF value below zero are no TR inhibitors. This is an acceptable failure rate; however, this false active rate can be minimized using a smaller range of values for active structures during the VS. On the basis of the PDDs observation, we selected only those structures with DF values between 0.5 and 2.0.
- The 739 promissory compounds selected through the VS procedure represent only 0.2% of the total screened structures. This (together with the fact that DF performed well in the external validation set of both active and non-active molecules) may indicate the DF generated has good selectivity, since we expect only a small number of compounds to be selected as promissory structures.

- 238 of the selected structures (32.2%) are nitro compounds. In contrast, the nitro group is present in only 0.3% of the total screened structures. This means the high percentage of nitro-compounds among the selected structures is derived from the DF classification and is not a consequence of a similar high percentage of nitro-compounds among the structures of our virtual screening library. Given the fact that there is strong presence (55.2%) of nitroaromatic drugs in the training set, the DF function is able to recognize the presence of the nitro function as one of the features that adds to the activity of the drug, which arises as further evidence of the chemical significance of the model. The high percentage of nitroaromatic compounds in the training set is a consequence of the fact that most identified TR inhibitors are nitro-compounds. However, there is still a high percentage of the selected structures that do not present the nitro-function in their molecular structure (67.8%), which indicates the inclusion of non-nitro drugs in the training set was a good strategy and that the DF is not biased towards the selection of nitro-compounds.
- 57.2% of the selected structures are fluorinated. If we consider that no fluorinated active drug was included in the training set and that further analysis of the 422,367 molecules analyzed in the VS has revealed that about 14% of the tested molecules contained fluorine, two remarks can be made:

  – The DF function is able to select molecules with structural features that do not appear in those of the training set, which means it can be used as a tool in the search of new leads. The high percentage of fluorinated drugs among the selected molecules is not a consequence of an analog high percentage of fluorinated drugs among the screened molecules.
  – From the greater incidence of fluorinated drugs within the 739 selected compounds compared to the incidence of fluorinated compounds within the 422,367 screened structures, it seems that apparently fluorine chemistry is a potential source of new antichagasic agents. This has been previously stated in literature and it seems to be confirmed by the present study [58]. Moreover, 31 of the 40 molecules with highest DF values are fluorinated (Figure 6a and b).

- 87 compounds (11.8%) among the 739 selected break the Lipinski rule of five, having one parameter out of range (the mlog P value). None of the 739 selected compounds has more than one parameter of the Lipinski rule of five out of range. This means all the selected structures are good candidates for oral administration.

Acknowledgments

We would like to thank Carolina Bellera and Rafael Franco for their valuable contribution to this paper. Alan Talevi is a fellowship holder of CONICET. This work was supported in part through grants from Agencia de Promoción Científica y Tecnológica (PICT 06-11985/2004) and Universidad Nacional de La Plata, Argentina.

References

1. Aguirre, G., Boiani, M., Cerecetto, H., Gerpe, A., González, M., Fernández – Sainz, Y., Denicola, A. and Escario, J.A., *Novel antiprotozoal products: imidazole and benzimidazole N-oxide derivatives and related compounds*, Arch. Pharm. (Weinheim), 337 (2004) 259–270.
2. Urbina, J.A., *Chemoterapy of Chagas Disease*, Current Pharmaceutical Design, 8 (2002) 287–295.

3. Faundez, M., Pino, L., Letelier, P., Ortiz, C., López, R., Seguel, C., Ferreira, J., Pavani, M., Morello, A. and Maya, J.D., *Buthionine Sulfoximine Increases the Toxicity of Nifurtimox and Benznidazole to Trypanosoma cruzi*, Antimicrobial Agents and Chemotherapy, 49 (2005) 126–130.
4. COST B9 Action on Chemotherapy of Protozoal Infections 1997–2002. Final report. European Cooperation in the Field of Scientific and Technical Research (COST). European Commission.
5. Urbina, J.A. and Docampo, R., *Specific chemotherapy of Chagas disease: controversies and advances*, Trends in Parasitology, 19 (2003) 495–501.
6. Andrade, S.G., Rassi, A., Magalhaes, J.B., Filho, F.F. and Luquetti, A.O., *Specific chemotherapy of Chagas disease: a comparison between the response in patients and experimental animals inoculated with the same strains*, Transactions of the Royal Society of Tropical Medicine and Higiene, 86 (1992) 624–626.
7. Oprea, T.I. and Matter, H., *Integrating Virtual Screening in Lead Discovery*, Current Opinion in Chemical Biology, 8 (2004) 349–358.
8. Drews, J., *Innovation Deficit Revisited: Reflections on the Productivity of Pharmaceutical R&D*, Drug Discovery Today, 3 (1998) 491–494.
9. Schneider, G. and Böhm, H., *Virtual Screening and Fast Automated Docking Methods*, Drug Discovery Today, 7 (2002) 64–70.
10. Lengauer, T., Lemmen, C., Carey, M. and Zimmermann, M., *Novel Technologies for Virtual Screening*, Drug Discovery Today, 9 (2004) 27–34.
11. Matter, H., *Selecting optimally diverse compounds from structure databases: A validation study of two-dimensional and three dimensional descriptors*, J. Med. Chem., 40 (1997) 1219–1229.
12. Matter, H. and Pötter, T., *Comparing 3D Pharmacophore Triplets and 2D Fingerprints for Selecting Diverse Compound Subsets*, J. Chem. Inf. Comput. Sci., 39 (1999) 1211–1225.
13. Patterson, D.E., Cramer, R.D., Ferguson, A.M., Clark, R.D. and Weinberger, L.E., *Neighborhood behavior: A useful concept for validation of "molecular diversity" descriptors*, J. Med. Chem., 39 (1996) 3049–3059.
14. Schuffenhauer, A., Gillet, V.J. and Willett, P., *Similarity Searching in Files of Three-Dimensional Chemical Structures: Analysis of the BIOSTER Database Using Two-Dimensional Fingerprints and Molecular Field Descriptors*, J. Chem. Inf. Comput. Sci., 40 (2000) 295–307.
15. Bajorath, J., *Selected Concepts and Investigations in Compound Classification, Molecular Descriptor Analysis, and Virtual Screening*, J. Chem. Inf. Comput. Sci., 41 (2001) 233–245.
16. Sheridan, R.P. and Kearsley, S.K., *Why do we need so many chemical similarity search methods*?, Drug Discovery Today, (2002) 903–911.
17. Devillers, J. and Balaban, A.T., *Topological Indices and Related Descriptors in QSAR and QSPR*, Gordon and Breach Science Publishers, (1999) 1–17.
18. Bonnet, B., Soullez, D., Girault, S., Maes, L., Landry, V., Davioud-Charvet, E. and Sergheraert, C., *Trypanothione Reductase Inhibition/Trypanocidal Activity Relationships in a 1,4-Bis(3-aminopropyl)piperazine Series*, Bioorg. and Med. Chem., 8 (2000) 95–103.
19. Castro-Pinto, D.B., Echevarria, A., Genestra, M.S., Cysne-Finkelstein, L. and Leon, L.L., *Trypanothione reductase activity is prominent in metacyclic promastigotes and axenic amastigotes of Leishmania amazonesis. Evaluation of its potential as a therapeutic target*, J. Enzyme Inhib. Med. Chem., 19 (2004) 57–63.
20. Meiering, S., Inhoff, O., Mies, J., Vincek, A., García, G., Kramer, B., Dormeyer, M. and Krauth-Siegel, R.L., *Inhibitors of Trypanosoma cruzi Trypanothione Reductase Revealed by Virtual Screening and Parallel Synthesis*, J. Med. Chem., 48 (2005) 4793–4802.
21. Cuellar, M.A., Salas, C., Cortés, M.J., Morillo, A., Maya, J.D. and Preite, M.D., *Synthesis and In Vitro Trypanocide Activity of Several Polycyclic Drimane-Quinone Derivatives*, Bioorg. and Med. Chem., 11 (2003) 2489–2497.
22. Carvalho, S.A., da Silva, E.F., Santa-Rita, R.M., de Castro, S.L. and Fraga, C.A.M., *Synthesis and antitrypanosomal profile of new functionalized 1,3,4-thiadiazole-2-arylhydrazone derivatives, designed as non-mutagenic megazol analogues*, Bioorg. and Med. Chem. Lett., 14 (2004) 5967–5970.
23. Machado, C.A. and Ayala, F.J., *Sequence variation in the dihydrofolate reductase-thymidylate synthase (DHFR-TS) and trypanothione reductase (TR) genes of Trypanosoma cruzi*, Molecular and Biochemical Parasitology, 121 (2002) 33–47.
24. Dragon Academic version is a product by Milano Chemometrics and QSAR research group, Milano, Italy.
25. Fernández, M., Caballero, J., Morales Helguera, A., Castro, E.A. and Pérez González, M., *Quantitative structure–activity relationship to predict differential inhibition of aldose reductase by flavonoid compounds*, Bioorg. and Med. Chem., 13 (2005) 3269–3277.
26. Gupta, M.K., Sagar, R., Shaw, A.K. and Prabhakar, Y.S., *CP-MLR directed QSAR studies on the antimycobacterial activity of functionalized alkenols–topological descriptors in modeling the activity*, Bioorg. and Med. Chem., 13 (2005) 343–351.
27. Puzyn, T. and Falandysz, J., *Computational estimation of logarithm of n-octanol/air partition coefficient and subcooled vapor pressures of 75 chloronaphthalene congeners*, Atmospheric Environment, 39 (2005) 1439–1446.
28. Garkani-Nejad, Z., Karlovits, M., Demuth, W., Stimpfl, T., Vycudilik, W., Jalali-Heravi, M. and Varmuza, K., *Prediction of gas chromatographic retention indices of a diverse set of toxicologically relevant compounds*, J. Chromatogr. A., 1028 (2004) 287–295.
29. StatSoft, Inc. (2004), Electronic Statistics Textbook. Tulsa, OK: Statsoft. WEB: http://www.statsoft.com/textbook/stathome.htlm.
30. Estrada, E., Uriarte, E., Montero, A., Teijeira, M., Santana, L. and De Clercq, E., *A novel approach for the virtual screening and rational design of anticancer compounds*, J Med Chem., 43 (2000) 1975–1985.
31. Diaz, H.G., Marrero, Y., Hernandez, I., Bastida, I., Tenorio, E., Nasco, O., Uriarte, E., Castanedo, N., Cabrera, M.A., Aguila, E., Marrero, O., Morales, A. and Perez, M., *3D-MEDNEs: an alternative 'in silico' technique for chemical research in toxicology. 1. prediction of chemically induced agranulocytosis*, Chem Res Toxicol, 16 (2003) 1318–1327.
32. Marrero Ponce, Y., Cabrera Perez, M.A., Romero Zaldivar, V., Gonzalez Diaz, H. and Torrens, F., *A new topological descriptors based model for predicting intestinal epithelial transport of drugs in Caco-2 cell culture*, J Pharm Pharm Sci., 7 (2004) 186–199.
33. Gálvez] García-García, A., Gálvez, J., de Julián-Ortiz, J.V., García-Domenech, R., Muñoz, C., Guna1, R. and Borrás, R., *New agents active against Mycobacterium avium complex selected by molecular topology: a virtual screening method*, Journal of Antimicrobial Chemotherapy, 53 (2004) 65–73.
34. Julián-Ortiz, J.V., Gálvez, J., Muñoz Collado, C., García-Domenech, R. and Gimeno-Cardona, C., *Virtual combinatorial syntheses and computational screening of new potential anti-herpes compounds*, J Med Chem., 42 (1999) 3308–3314.
35. Meneses-Marcel, A., et al., *A linear discriminant analysis based virtual screening of trichomonacidal lead – like compounds: Outcomes of in silico studies supported by experimental results*, Bioorganic and Medicinal Chemistry Letters, 15 (2005) 3838–3843.
36. The Merck Index., Encyclopedia Of Chemicals, Drugs, And Biologicals, 13th ed. Merck & Co. Whitehouse Station, NJ, 2001.
37. Salmon-Chemin, L., Buisine, E., Yardley, V., Kohler, S., Debreu, M.A., Landry, V., Sergheraert, C., Croft, S.L. and Krauth-Siegel, R.L., *Davioud-Charvet E. 2- and 3-substituted 1,4-naphthoquinone derivatives as subversive substrates of trypanothione reductase and lipoamide dehydrogenase from Trypanosoma cruzi: synthesis and correlation between redox cycling activities and in vitro cytotoxicity*, J. Med. Chem., 44 (2001) 548–565.
38. Mayaa, J.D., Bollob, S., Nuñez-Vergara, L.J., Squellab, J.A., Repetto, Y., Morello, A., Périéc, J. and Chauvière, G., *Trypanosoma cruzi: effect and mode of action of nitroimidazole and nitrofuran derivatives*, Biochemical Pharmacology, 65 (2003) 999–1006.

39. Kinnamon, K.E., Poon, B.T., Hanson, W.L. and Waits, V.B., *Trypanosoma cruzi: A Novel Chemical Class (Nitrobenzofurans) Active against Infections of Mice (Mus musculus)*, Experimental Parasitology, 89 (1998) 251–256.
40. Dixon, W.J., *BMDP Biomedical Computer Programs*. Version 2.0. University of California, Los Angeles. 2001.
41. Statsoft, Inc. STATISTICA. Version 5.1. Tulsa, OK: Statsoft., 1996.
42. Gálvez, R., García-Domenech, C., Alapont de, G., Julián-Ortiz, J.V. and Popa, L., *Pharmacological distribution diagrams: A tool for de novo drug design*, Journal of Molecular Graphics, 14 (1996) 272–276.
43. Moriguchi, I., Hirono, S., Nakagome, I. and Hirano, H., *Comparison of Reliability of logP Values for Drugs Calculated by Several Methods*, Chem. Pharm. Bull, 42 (1994) 976–978.
44. Daunes, S. and D'Silva, C., *QSAR Study on the Contribution of Log P and Es to the in Vitro Antiprotozoal Activity of Glutathione Derivatives*, J. Med. Chem., 44 (2001) 2976–2983.
45. Daunes, S. and D'Silva, C., *Glutathione Derivatives Active against Trypanosoma brucei rhodesiense and T. brucei brucei In Vitro*, Antimicrobial Agents and Chemotherapy, 46 (2002) 434–437.
46. Aguirre, G., et al., *Design, synthesis and biological evaluation of new potent 5-nitrofuryl derivatives as anti-Trypanosoma cruzi agents. Studies of trypanothione binding site of trypanothione reductase as target for rational design*, European Journal of Medicinal Chemistry, 39 (2004) 421–431.
47. Turner, J., Glass, B.D. and Agatonovic-Kustrin, S., *Prediction of drug bioavailability based on molecular structure*, Analytica Chimica Acta, 485 (2003) 89–102.
48. Ghose, A.K., Viswanadhan, V.N. and Wendoloski, J.J., *A Knowledge-Based Approach in Designing Combinatorial or Medicinal Chemistry Libraries for Drug Discovery. A Qualitative and Quantitative Characterization of Known Drug Databases*, J. Comb. Chem., 1 (1999) 55–68.
49. Smith, C.J. and Hansch, C., *The Relative Toxicity of Compounds in Mainstream Cigarette Smoke Condensate*, Food and Chemical Toxicology, 38 (2000) 637–646.
50. Proudfoot, J.R., *The evolution of synthetic oral drug properties*, Bioorg. and Med. Chem. Lett., 15 (2005) 1087–1090.
51. Pouton, C.W., *Lipid formulations for oral administration of drugs: non-emulsifying, self-emulsifying and 'self-microemulsifying' drug delivery systems*, European Journal of Pharmaceutical Sciences, 11(2) (2002) S93–S98.
52. Xue, C.X., Zhang, X.I., Liu, M.C., Hu, Z.D. and Fan, B.T., *Study of probabilistic neural networks to classify the active compounds in medicinal plants*, Journal of Pharmaceutical and Biomedical Analysis, 38 (2005) 497–507.
53. Gálvez, J., García, R., Salabert, M.T. and Soler, R., *Charge Indexes. New Topological Descriptors*, J. Chem. Inf. Comput. Sci., 34 (1994) 520–525.
54. Silverman, B.D., *Three-Dimensional Moments of Molecular Property Fields*, J. Chem. Inf. Comput. Sci., 40 (2000) 1470–1476.
55. Bath, P.A., Poirrette, A.R. and Willett, P., *The Extent of the Relationship between the Graph-Theoretical and the Geometrical Shape Coefficients of Chemical Compounds*, J. Chem. Inf Comput. Sci., 35 (1995) 714–716.
56. Milicevic, A., Nikolic, S. and Trinajstic, N., *On reformulated Zagreb indices*, Molecular Diversity, 8 (2004) 393–399.
57. Ismail, F.M.D., *Important fluorinated drugs in experimental and clinical use*, Journal of Fluorine Chemistry, 118 (2002) 27–33.