

# Computational Discovery of Novel Trypanosomicidal Drug-like Chemicals by Using Bond-based Non-stochastic and Stochastic Quadratic Maps and Linear Discriminant Analysis.

Juan Alberto Castillo-Garit,<sup>a,b,c\*</sup> Maria C. Vega,<sup>d</sup> Miriam Rolon,<sup>d</sup> Yovani Marrero-Ponce,<sup>b,c,f</sup> Vladimir V. Kouznetsov,<sup>g</sup> Diego Fernando Amado Torres,<sup>g</sup> Alicia Gómez-Barrio,<sup>d</sup> Alfredo Alvarez Bello,<sup>b</sup> Alina Montero,<sup>b</sup> Francisco Torrens,<sup>c</sup> and Facundo Pérez-Giménez.<sup>f</sup>

<sup>a</sup>Applied Chemistry Research Center, Faculty of Chemistry-Pharmacy, Central University of Las Villas, Santa Clara, 54830, Villa Clara, Cuba. e-mail: [jacgarit@yahoo.es](mailto:jacgarit@yahoo.es), [juancg.22@gmail.com](mailto:juancg.22@gmail.com) or [juancg@uclv.edu.cu](mailto:juancg@uclv.edu.cu)

<sup>b</sup>Unit of Computer-Aided Molecular "Biosilico" Discovery and Bioinformatic Research (CAMD-BIR Unit), Faculty of Chemistry-Pharmacy, Central University of Las Villas, Santa Clara, 54830, Villa Clara, Cuba.

<sup>c</sup>Institut Universitari de Ciència Molecular, Universitat de València, Edifici d'Instituts de Paterna, P.O. Box 22085, E-46071, València, Spain.

<sup>d</sup>Departamento de Parasitología, Facultad de Farmacia, UCM, Pza. Ramón y Cajal s/n, 28040 Madrid.

<sup>f</sup>Unidad de Investigación de Diseño de Fármacos y Conectividad Molecular, Departamento de Química Física, Facultad de Farmacia, Universitat de València, Spain.

<sup>g</sup>Laboratorio de Química Orgánica y Biomolecular, Escuela de Química, Universidad Industrial de Santander, Bucaramanga, Colombia.

---

## Abstract

Herein we present results of a Quantitative Structure-Activity Relationship (QSAR) studies to classify and design, in a rational way, new antitrypanosomal compounds by using non-stochastic and stochastic bond-based quadratic indices. A data set of 440 organic chemicals, 143 with antitrypanosomal activity and 297 having other clinical uses, is used to develop QSAR models based on Linear Discriminant Analysis (LDA). Non-stochastic model correctly classifies more than 93% and 95% of chemicals in both training and external prediction groups, respectively. On the other hand, the stochastic model shows an accuracy of about the 87% for both series. As an experiment of virtual lead generation, the present approach is finally satisfactorily applied to the virtual evaluation of 9 already synthesized *in house* compounds. The *in vitro* antitrypanosomal activity of this series against epimastigote forms of *Trypanosoma cruzi* is assayed. The model is able to predict correctly the behaviour for the majority of these compounds. Four compounds (FER16, FER32, FER33 and FER 132) showed more than 70% of epimastigote inhibition at a concentration of 100µg/mL (86.74%, 78.12%, 88.85% and 72.10%, respectively) and two of these chemicals, FER16 (78.22% of AE) and FER33 (81.31% of AE), also showed good activity at a concentration of 10µg/mL. At the same concentration, compound FER16 showed lower value of cytotoxicity (15.44%), and compound FER33 showed very low value of 1.37%. Taking into account all these results, we can say that these three compounds can be optimized in forthcoming works, but we consider that compound FER33 is the best candidate. Even though none of them resulted more active than Nifurtimox, the current results constitute a step forward in the search for efficient ways to discover new lead antitrypanosomals.

**Keywords:** TOMOCOMD-CARDD Software, Bond-based Quadratic Indices, LDA-assisted QSAR Model, virtual Screening, Trypanosomicidal, Cytotoxicity.

## 1. Introduction

Parasitic diseases affect hundreds of millions of people worldwide and result in significant mortality and devastating social and economic consequences. Nevertheless, most of the drugs available to treat these diseases are decades old and are frequently limited in efficacy, plagued by severe side effects and poor patient compliance, or hamstrung by drug resistance. Few, if any, of the currently available drugs for parasitic diseases would pass through even a discovery-stage screening funnel today, letting apart preclinical and clinical development (Renslo and McKerrow, 2006). The current state of chemotherapeutics for parasitic diseases is particularly bleak for those living in the affected regions of the world because of the low economic incentive for drug development and the rise of resistant strains (Weisman et al., 2006). These diseases, though globally massive in their impact, affect mainly poor people in undeveloped regions of the world. As such, they would never be viewed as viable target markets for the pharmaceutical industry, particularly in today's post-merger climate. In parallel, funding for basic research on these organisms and the pathogenesis of the diseases they produce has been woefully inadequate compared with funding for diseases of much lower prevalence but more direct impact in the developed countries of Europe and North America (Renslo and McKerrow, 2006). However, in recent times the Chagas' disease is appearing also in countries of the First World due to of immigration, organ donation and blood transfusion.

Among the parasitic-diseases, protozoa are responsible for a number of illnesses, including leishmaniasis, Chagas disease, malaria, schistosomiasis, African trypanosomiasis and giardiasis/amebiasis. In this sense, Chagas disease or American trypanosomiasis (caused by *Trypanosoma cruzi*) occupies the third place in the number of deaths per year, after malaria and schistosomiasis (Aguirre et al., 2004). 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 (Prieto et al., 2006). Current chemotherapy against *Trypanosoma* remains unsatisfactory; available drugs are benznidazole and nitrofurans such as nifurtimox. The latter has undergone several rumors of discontinuation (Urbina, 2002; Faundez et al., 2005), probably because of limited markets or the potential risks the pharmaceutical companies may incur because of suspicion of long-term toxicity (COSTB9, 1997–2002). Both drugs have significant activity in only 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 (Urbina

and Docampo, 2003). Their efficacy also varies according to geographical areas, mainly because of differences in drug susceptibility of different *T. cruzi* strains (Andrade et al., 1992; Urbina, 2002).

Screening large chemical libraries to identify compounds with trypanocidal activity has been hindered in the past by the lack of efficient screening assays with available assays being labour intensive, relying on expensive instrumentation, or requiring radioisotopes (Kaminsky and Brun, 1993). 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 (Prieto et al., 2006). 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 *CO*mputer *DE*sign] (Marrero-Ponce and Romero, 2002; Marrero-Ponce, 2003; Marrero Ponce, 2004; Marrero-Ponce et al., 2006c; Casanola-Martin et al., 2007; Marrero-Ponce et al., 2007; Marrero-Ponce et al., 2008) allows us to perform rational *in silico* molecular design (selection/identification) and Quantitative Structure-Activity/Property Relationship (QSAR/ QSPR) studies. Therefore, this scheme has been applied to the prediction of several physical, physicochemical, chemical, pharmacokinetical, toxicological as well as biological properties (Marrero-Ponce, 2004; Marrero-Ponce et al., 2004; Marrero-Ponce et al., 2005a; Casañola-Martin et al., 2006; Marrero-Ponce et al., 2006b; Marrero-Ponce et al., 2007; Castillo-Garit et al., 2008a; Castillo-Garit et al., 2008b). 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* (Marrero-Ponce et al., 2005b). Other studies for the rational discovery of novel paramphistomocides (Marrero-Ponce et al., 2005c), antimalarial (Marrero-Ponce et al., 2005d) and antibacterial (Marrero-Ponce et al., 2006a) compounds were also conducted with the *TOMOCOMD* approach; also studies related to proteomics (Marrero-Ponce et al., 2005e) and nucleic acid-drug (Marrero Ponce et al., 2005) interactions have been carried out. In addition, 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 (Marrero-Ponce and Castillo-Garit, 2005; Castillo-Garit et al., 2006; Castillo-Garit et al., 2007; Castillo-Garit et al., 2008c).

In the present report, bond-based non-stochastic and stochastic quadratic indices are used to find classification models that allow the discrimination of antitrypanosomal compounds. This kind of approach permits the rational identification of those candidates to be evaluated, which have the

highest probabilities of being active ones. Following this idea, eight already-synthesized compounds were then *in silico* evaluated and, after that, *in vitro* assayed against epimastigote forms of *Trypanosoma cruzi*. Cytotoxic studies were also conducted, as selection criterion of compounds to be evaluated in further anti-amastigote and *in vivo* assays.

## 2. Materials and Methods

### 2.1. Data set for QSAR Study

The general data set used in this study consists of 440 compounds of great structural variation, 143 of which are actives and 297 are inactive against trypanosome. The antitrypanosomals considered in this study are representative of families with diverse structural patterns and were collected from previous publications (Gillmor et al., 1997; Urbina et al., 1998; Bonse et al., 1999; Cerecetto et al., 2000; Hiyoshi et al., 2000; Werbovetz, 2000; Buckner et al., 2001; Salmon-Chemin et al., 2001; Zuccotto et al., 2001; Daunes and D'Silva, 2002; Du et al., 2002; Elhalem et al., 2002; Gilbert, 2002; Bal et al., 2003; Buckner et al., 2003; Hamilton et al., 2003; Huang et al., 2003; Urbina et al., 2004). The names of compounds in the database together with their experimental data taken from the literature are reported in the Supporting Information (activity data and structures of the 143 anti-trypanosome agents in Tables S1 and Figures S1, respectively). It is remarkable that the wide variability of drugs and mechanisms of action of active compounds in the training and prediction sets assures adequate extrapolation power and increases the possibilities of the discovery of new lead compounds with novel mechanisms of action, which results one of the most critical aspects in the construction of non-congeneric data. Therefore, this dataset provide us with a suitable data matrix for investigating the potential of computational tools in ligand-based drug design of trypanocidal agents

On the other hand, 297 compounds having different clinical uses such as antivirals, sedative/hypnotics, diuretics, anticonvulsants, hemostatics, oral hypoglycemics, anti-hypertensives, antihelminthics and anticancer compounds as well as some other kinds of drugs were selected for the set of inactive compounds through random selection, guaranteeing great structural variability as well. All these compounds were taken from the Negwer Handbook (Negwer, 1987) and Merck Index (1996) in which their names, synonyms, and structural formulas can be found. The classification of these organic compounds as 'inactive' (non-antitrypanosomal) does not guarantee that all are truly so; some of them may have inhibitory activity toward tyrosinase, which is undetected. This problem can be reflected in the results of classification for the series of inactive compounds (Estrada and Peña, 2000).

## 2.2. Computational approach

The theory of the bond-based quadratic indices used in this study was discussed in detail in an earlier publications (Marrero-Ponce et al., 2006c). Specifically, the *CARDD* (Computed-Aided Rational Drug Design) module implemented in the *TOMOCOMD* Software (Marrero-Ponce and Romero, 2002) was used in the calculation of bond-based non-stochastic and stochastic quadratic indices. In this study, the properties used to differentiate the molecular atoms are those previously proposed for the calculation of the DRAGON descriptors (Kier and Hall, 1986; Todeschini and Gramatica, 1998; Consonni et al., 2002), i.e., atomic mass (M), atomic polarizability (P), atomic Mulliken electronegativity (K), van der Waals atomic volume (V), plus the atomic electronegativity in Pauling scale (G)(Pauling, 1939).

The bond-based *quadratic indices* descriptors computed in this study were the following:

- 1)  $k^{\text{th}}$  ( $k = 15$ ) total non-stochastic bond-based quadratic indices, not considering and considering H-atoms in the molecular graph (G) [ $q_k(\bar{w})$  and  $q_k^{\text{H}}(\bar{w})$ , respectively].
- 2)  $k^{\text{th}}$  ( $k = 15$ ) total stochastic bond-based quadratic indices, not considering and considering H-atoms in the molecular graph (G) [ ${}^s q_k(\bar{w})$  and  ${}^s q_k^{\text{H}}(\bar{w})$ , respectively].
- 3)  $k^{\text{th}}$  ( $k = 15$ ) bond-type local (group = heteroatoms: S, N, O) non-stochastic quadratic indices, not considering and considering H-atoms in the molecular graph (G) [ $q_{kL}(\bar{w}_E)$  and  $q_{kL}^{\text{H}}(\bar{w}_E)$ , correspondingly]. These local descriptors are putative molecular charge, dipole moment, and H-bonding acceptors.
- 4)  $k^{\text{th}}$  ( $k = 15$ ) bond-type local (group = heteroatoms: S, N, O) stochastic quadratic indices, not considering and considering H-atoms in the molecular graph (G) [ ${}^s q_{kL}(\bar{w}_E)$  and  ${}^s q_{kL}^{\text{H}}(\bar{w}_E)$ , correspondingly]. These local descriptors are also putative molecular charge, dipole moment, and H-bonding acceptors.

## 2.3. Chemometric method

Linear discriminant analysis (LDA) was performed with software package STATISTICA (STATISTICA version. 6.0, 2001). Forward stepwise was fixed as the strategy for variable selection. The quality of the models was determined by examining Wilk's  $\lambda$  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) (Baldi et al., 2000). 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 by using an external prediction (test) set.

## 2.4 Biological Assay: Determination of 'in vitro' Tripanosomicidal Activity and Cytotoxicity

### 2.4.1 Parasites and culture procedure

The strain Y of *T. cruzi* (Silva and Nussensweig, 1953) was originally isolated from an acute human case coming from Marília (São Paulo, Brazil) in 1950. Epimastigotes were grown at 28° C in liver infusion tryptose broth (LIT) with 10% fetal bovine serum (FBS), penicillin and streptomycin as previously described (Gómez-Barrio et al., 1997).

### 2.4.2 Epimastigotes susceptibility assay

The activity was evaluated with resazurin by colorimetric method described previously (Rolón et al., 2006b). The screening assay was performed in 96-well microplates with cultures in LIT with 10% FBS, which had not reached the stationary phase. Epimastigotes were seeded at  $3 \times 10^6$  per milliliter in culture tubes. Following a 24 h incubation to allow homogeneous growth, 200 µl volumes were seeded in the plates in the presence of serial dilutions of reference drugs (concentration range as above) at 28° C for 48 hours, at which time 20 µl of resazurin solution 3mM was added and returned to the incubator for another 5 h. A solution of resazurin was prepared in 1% phosphate buffer solution (PBS), pH 7 and filter sterilized before use. Growth controls were also included. The oxidation-reduction was quantitated at 490 and 595 nm. Each concentration was assayed in triplicate. In order to avoid drawback, medium and drug controls were used in each test. The anti-epimastigotes percentage (%AE) was calculated as follows:

$$\%AE = [(ALW - (AHW \times RO) \text{ test well}) / (ALW - (AHW \times RO) \text{ positive growth control})] \times 100$$

where, ALW and AHW represents the absorbances at the lower and the higher wavelength respectively (medium was subtracted) and RO represents the correction factor ( $RO = ALW/AHW$  for resazurin in medium).

### 2.4.3 Cell culture

The cell lines used were National Collection of Type Cultures (NCTC) clone 929 and murine J774 macrophages. NCTC clone 929 cells were grown in Minimal Essential Medium (Sigma) and J774 macrophages were grown in RPMI 1640 medium (Sigma). Both media were supplemented with 10% heat-inactivated FBS (30 minutes at 56°C), penicillin G (100 U/mL) and streptomycin (100 µg/mL). For the experiments, cells in the pre-confluence phase were harvested with trypsin. Cell cultures were maintained at 37°C in a humidified 5% CO<sub>2</sub> atmosphere.

#### 2.4.4 Cytotoxicity assays

The procedure for cell viability measurement was evaluated with resazurin by a colorimetric method described previously (Rolón et al., 2006a; Rolón et al., 2006b). The macrophages J774 were seeded ( $5 \times 10^4$  cells/well) in 96-well flat-bottom microplates with 100  $\mu$ l of RPMI 1640 medium. The cells were allowed to attach for 24 h at 37°C, 5% CO<sub>2</sub> and the medium was replaced by different concentrations of the drugs in 200  $\mu$ l of medium, and exposed for another 24 h. Growth controls were also included. Afterwards, a volume 20  $\mu$ l the 2mM resazurin solution was added and plates were returned to incubator for another 3 h. to evaluate cell viability. The reduction of resazurin was determined by dual wavelength absorbance measurement at 490 nm and 595 nm. Background was subtracted. Each concentration was assayed in triplicate. Medium and drug controls were used as blanks in each test.

### 3. Results and Discussion

#### 3.1. LDA-QSAR Models: Developing and validation.

The LDA has become an important tool for the prediction of chemical properties. Due to the simplicity of this method many useful discriminant models have been developed and presented by different authors in the literature (Estrada and Peña, 2000; Estrada et al., 2000; Marrero Ponce et al., 2005; Casañola-Martin et al., 2006; Marrero-Ponce et al., 2006b; Castillo-Garit et al., 2008b; Castillo Garit et al., 2008). It was the technique used in the generation of a discriminant function in the present work. The principle of maximal parsimony (Occam's razor) was taken into account as the strategy for model selection (Estrada, 1999). The general data set was randomly divided into two subsets, training and test set (which have 346 and 94 compounds, respectively), both of them containing active and inactive compounds. The best models obtained using bond-based non-stochastic and stochastic quadratic indices as molecular descriptors, together with their statistical parameters are given below, respectively:

$$\text{Class} = -2.77 - 1.73 \times 10^{-1} G_{0L}^H(\bar{x}_E) + 8.62 \times 10^{-2} G_{1L}^H(\bar{x}) - 5.94 \times 10^{-4} G_{4L}^H(\bar{x}) \\ + 7.10 \times 10^{-1} G_{12L}^H(\bar{x}_E) - 2.32 \times 10^{-2} G_{11L}^H(\bar{x}_E) - 7.52 \times 10^{-7} G_{8L}^H(\bar{x}) \quad (1)$$

$$N = 346 \quad \lambda = 0.35 \quad Q_{Total} = 93.35 \% \quad MCC = 0.86$$

$$D^2 = 8.10 \quad F = 104.32 \quad p < 0.0001$$

$$\text{Class} = -3.54 - 1.04 \times 10^{-2} M_{2L}(\bar{x}) + 5.27 \times 10^{-3} M_{14L}(\bar{x}_E) + 1.54 \times 10^{-2} M_{4L}(\bar{x}) \\ + 5.43 \times 10^{-3} M_{2L}(\bar{x}_E) + 1.44 \times 10^{-2} M_{0L}^H(\bar{x}) - 1.40 \times 10^{-2} M_{0L}^H(\bar{x}_E)$$

$$\begin{aligned}
 & - 8.01 \times 10^{-3} \text{ Ms } \mathbf{q}_3(\bar{x}) + 4.35 \times 10^{-2} \text{ Ms } \mathbf{q}_{15L}(\bar{x}_E) - 1.43 \times 10^{-1} \text{ Ms } \mathbf{q}_{11L}(\bar{x}_E) \\
 & + 9.13 \times 10^{-2} \text{ Ms } \mathbf{q}_{9L}(\bar{x}_E)
 \end{aligned} \tag{2}$$

$$\begin{aligned}
 N = 346 \quad \lambda = 0.50 \quad Q_{Total} = 87.57 \% \quad MCC = 0.73 \\
 D^2 = 4.41 \quad F = 33.69 \quad p < 0.0001
 \end{aligned}$$

where, N is the number of compounds,  $\lambda$  is the Wilks' statistic,  $Q_{Total}$  is the accuracy of the model for the training set, MCC is the Matthews' correlation coefficient,  $D^2$  is the square Mahalanobis distance, F is the Fisher ratio and  $p$ -value is its significance level.

The non-stochastic model (Eq. 1), which includes non-stochastic indices, has an accuracy of 93.35% for the training set. This model showed a high MCC of 0.86; MCC quantifies the strength of the linear relation between the molecular descriptors and the classifications, and it may usually provide a much more balanced evaluation of the prediction than, for instance, the percentages (accuracy) (Baldi et al., 2000). Nevertheless, the most important criterion, for the acceptance or not of a discriminant model, is based on the statistics for external prediction set. The non-stochastic model showed an accuracy of 95.74% (MCC = 0.90) for the compounds in the test set.

On the other hand, a stochastic linear indices model was obtained (Eq. 2), this model achieved an accuracy of 87.57% with a MCC of 0.73, for the test set the results of this model were an accuracy of 86.17% and MCC of 0.69; these values are acceptable, but lower than those obtained with non-stochastic quadratic indices. These results are given in Table 1.

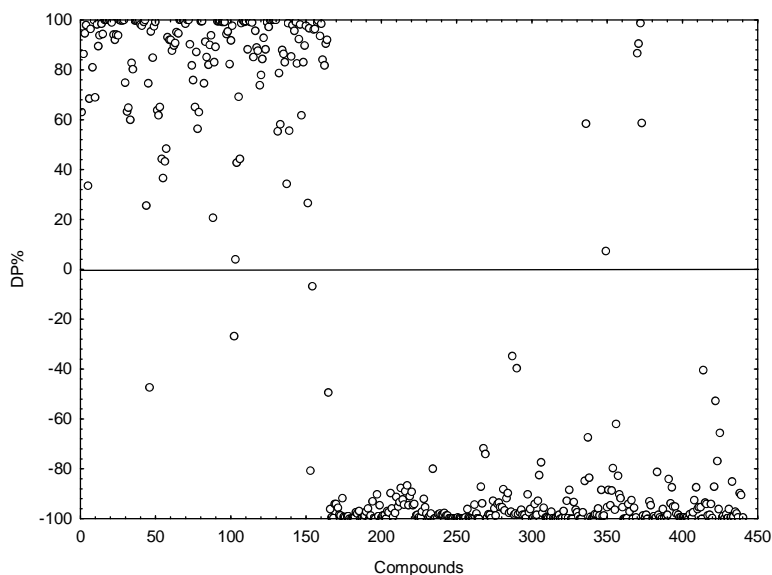
**Table 1.** Prediction performances for LDA-based QSAR models for training and test sets.

Models <sup>a</sup>	Matthews Corr. Coefficient (C)	Accuracy 'Q <sub>Total</sub> ' (%)	Specificity (%)	Sensitivity 'hit rate' (%)
Training set				
Eq. 1	0.86	93.35	84.89	98.33
Eq. 2	0.73	87.57	78.95	87.50
Test set				
Eq. 1	90.00	95.74	85.19	100.00
Eq. 2	69.00	86.17	65.63	91.30

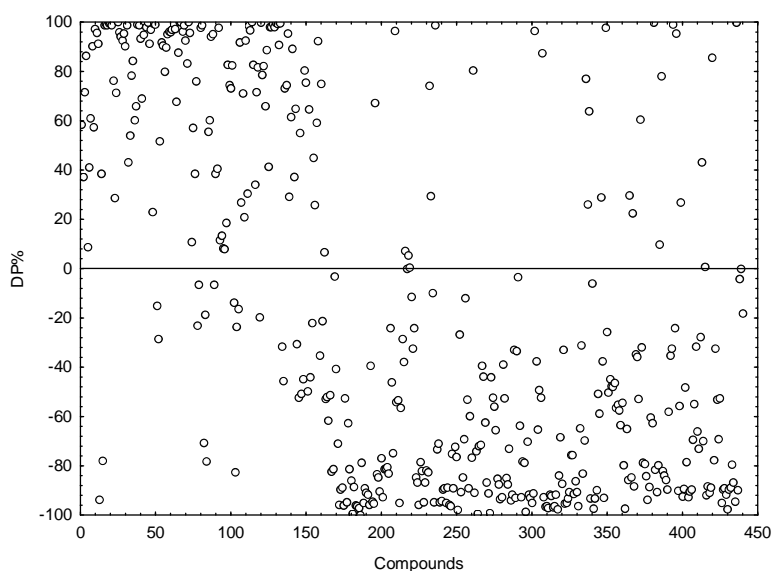
In addition, the probability of correctly predicting a positive example (sensitivity or hit rate) and the probability that a positive prediction will be correct (specificity) were computed for both models. The values obtained for sensibility were 98.33% and 87.50% for non-stochastic and stochastic models, correspondingly. While these 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, values of specificity were 84.89% and 78.95% for Eqs 1 and 2, correspondingly. Moreover, Figures 1 and 2 give a plot of the  $\Delta P\%$  for



the classification of all compounds in both training and test sets from models **1** and **2**, correspondingly.



**Figure 1.** Plot of the  $\Delta P\%$  from Eq. 1 (using non-stochastic quadratic indices) for each compound in the training and test sets. Compounds 1–120 and 121–143 are active (antitrypanosomal) in training and test sets, respectively; chemicals 144–345 and 346–440 are inactive (non-antitrypanosomal) in both training and test sets, correspondingly.



**Figure 2.** Plot of the  $\Delta P\%$  from Eq. 2 (using stochastic quadratic indices) for each compound in the training and test sets. Compounds 1–120 and 121–143 are active (antitrypanosomal) in training and test sets, respectively; chemicals 144–345 and 346–440 are inactive (non-antitrypanosomal) in both training and test sets, correspondingly.

### 3.2. Lead generation by using virtual screening. Identification and Experimental proof.

The performance of the results obtained above encouraged us to carry out an *in silico* screening to search for novel lead compounds with antitrypanosomal activity, as a way to show the applicability of the QSAR models obtained with the **TOMOCOMD-CARDD** approach, in the selection of *hit* (or lead) compounds. In order to find promising active agents, we selected a pool of compounds not yet described in the literature as trypanocidals. Later the *in silico* essays were performed by using all the models developed inside this report, to find bioactive chemicals that present trypanocidal activity.

Here, nine new organic compounds were evaluated with the LDA-based QSAR models, and the *in vitro* assays of the synthesized compounds were carried out to corroborate the *in silico* predictions. We proceeded to test the compounds in an epimastigote inhibition (*in vitro*) assay (Vega et al., 2005). The  $\Delta P\%$  values of the compounds in the data using all the discriminant functions and the chemical structures are depicted in Table 2 and Figure 1. A good agreement is observed between the experimental antitrypanosomal activity and theoretical predictions for most of the compounds. Four compounds (FER16, FER32, FER33 and FER 132) showed more than 70% of epimastigote inhibition at a concentration of 100 $\mu\text{g}/\text{mL}$  (86.74%, 78.12%, 88.85% and 72.10%, respectively); in addition, compound FER19 showed a value of %AE = 69.19% very similar to the cut value (70% of AE). Two compounds, FER16 (78.22% of AE) and FER33 (81.31% of AE), also showed good activity at a concentration of 10 $\mu\text{g}/\text{mL}$ . Even though none of them resulted more active than Nifurtimox, the current results constitute a step forward in the search for efficient ways to discover new lead antitrypanosomals. The remaining compounds which were classified as inactive for the model, showed very low inhibition percentages.

After this preliminary *in vitro* test, the unspecific cytotoxicity was determined against macrophages at the concentrations that were used in the previous assay (Rolón et al., 2006a; Rolón et al., 2006b). At this time, two (FER16, and FER33) of the four compound that shown more than 70% of epimastigote inhibition at a concentration of 100 $\mu\text{g}/\text{mL}$  also showed high values of cytotoxicity (53.20% and 44.06%, respectively). At a concentration of 10 $\mu\text{g}/\text{mL}$ , compound FER16 showed lower value of cytotoxicity, (15.44%) and compound FER33 showed very low value of 1.37%. On the other hand, the compound FER32 showed only 11.29% of cytotoxicity at a concentration of 100 $\mu\text{g}/\text{mL}$ . Taking into account all these results, we can say that these three compounds can be optimized in forthcoming works, but we consider that compound FER33 is the best candidate.

**Table 2.** Compounds evaluated in the present study, their classification ( $\Delta P\%$ ) according to the obtained models, their antitrypanosomal activity and cytotoxicity at three different concentrations (100, 10, and 1  $\mu\text{g/mL}$ ) and antitrypanosomal activity of nifurtimox (reference).

Compound	Exp. <sup>a</sup>	$\Delta P$ Eq. 1 <sup>b</sup>	$\Delta P$ Eq. 2 <sup>c</sup>	%AE (SD) <sup>d</sup>			%CI <sup>e</sup>		
				100 $\mu\text{g/mL}$	10 $\mu\text{g/mL}$	1 $\mu\text{g/mL}$	100 $\mu\text{g/mL}$	10 $\mu\text{g/mL}$	1 $\mu\text{g/mL}$
FER10	I	-80.03	-3.17	15.32 $\pm$ 1,4	5.07 $\pm$ 1,28	0.00 $\pm$ 2,10	21,275	7,65	0
FER16	A	-22.93	62.58	<b>86.74<math>\pm</math> 1,73</b>	<b>78.22<math>\pm</math> 1,37</b>	19.85 $\pm$ 2,01	53,02	<b>15,44</b>	4,43
FER19	A	-7.13	69.25	69.19 $\pm$ 1,02	3.17 $\pm$ 0,28	2.93 $\pm$ 2,90	97,78	12,82	6,11
FER25	I	-79.51	-4.10	59.13 $\pm$ 0,53	42.87 $\pm$ 1,4	39.31 $\pm$ 2,66	20,83	6,47	9,188
FER26	I	-72.76	1.04	10.39 $\pm$ 2,05	4.10 $\pm$ 2,21	0.84 $\pm$ 1,68	6,11	2,62	1,88
FER29	I	-99.67	86.76	13.10 $\pm$ 1,22	1.98 $\pm$ 1,48	0.00 $\pm$ 1,13	37,25	7,11	6,04
FER32	A	21.67	71.30	<b>78.12<math>\pm</math> 0,71</b>	44.99 $\pm$ 2,52	4.55 $\pm$ 2,44	<b>11.29</b>	0	0
FER33	A	19.37	71.68	<b>88.85<math>\pm</math> 3,44</b>	<b>81.31<math>\pm</math> 0,76</b>	39.03 $\pm$ 4,34	44,06	<b>1,37</b>	0
FER132	A	-89.49	53.82	<b>72.10<math>\pm</math>0.28</b>	38.20 $\pm$ 2.61	14.83 $\pm$ 5.16	100	100	94,95
Nifurtimox	A	99.98	98.39	<b>100<math>\pm</math>1.49</b>	<b>85.45<math>\pm</math>2.43</b>	38.21 $\pm$ 2.17	11.68	0.6	0.32

<sup>a</sup>Observed activity.

<sup>b</sup>Results of the classification of compounds obtained from Model 1,  $DP\% = [P(\text{active}) - P(\text{inactive})] \cdot 100$

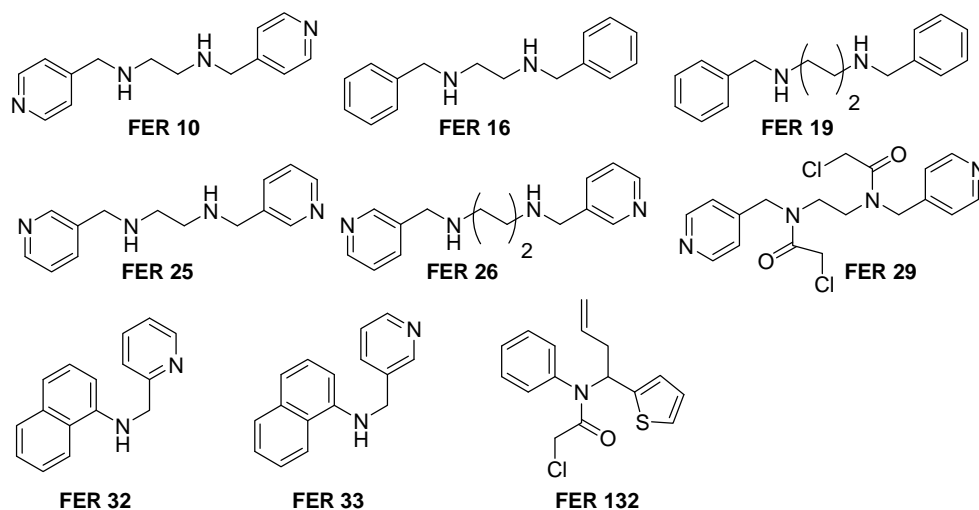
<sup>c</sup>Results of the classification of compounds obtained from Model 2,  $DP\% = [P(\text{active}) - P(\text{inactive})] \cdot 100$

<sup>d</sup>Anti-epimastigotes percentage and standard deviation (SD)

<sup>e</sup>Cytotoxicity percentage

A: active

I: inactive



**Figure 3.** Structures of the experimentally evaluated compounds.

#### 4. Conclusions

The research involving the discovery of new trypanosomicidals is considered an impacting field in pharmaceutical and therapeutical areas. This fact is due to the fact that Chagas' disease occupies the third place in the number of deaths per year in Latin America. The great number of people infected with *T. cruzi* and the millions in risk of being infected and the low efficacy of the actual treatments make of this disease one of the major health problems in Latin America. However, the discovery research activities are in general extremely time-consuming and expensive; therefore, it is imperative to develop novel alternative techniques. Moreover, the use of *in silico* approaches has emerged as a replacement alternative to *in vivo* test assays. In spite of some criticism, topological indices-based approaches have demonstrated their usefulness in drug discovery processes.

**TOMOCOMD-CARDD** method 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*. In the present report, the usefulness of the non-stochastic and stochastic bond-based quadratic indices was shown to discriminate antitrypanosomal compounds from inactive ones and to predict the activity of novel compounds against *T. cruzi*. Furthermore, four out of nine new compounds, subjected to *in silico* screening, were recognized with antitrypanosomal activity. Afterward, several *in vitro*

experiments were performed to corroborate the reliability of the classification functions developed in this work and permit us to select the candidates with the best “activity against epimastigote forms/unspecific cytotoxicity” rate.

The interactive and flexible character of the **CARDD** scheme permits the posterior inclusion of other active and inactive compounds in the training set and the generation, at each step, of more refined models capable of identifying structural patterns not considered in the present study. Finally, we can say that bond-based quadratic indices can be successfully used in the future for the rational search for novel antitrypanosomal compounds.

**Acknowledgement:** Castillo-Garit, J.A. and Marrero-Ponce, Y.; thanks the program ‘Estades Temporals per a Investigadors Convidats’ for a fellowship to work at Valencia University in 2008 and 2009, respectively. Kouznetsov, V.V. thank the Instituto Colombiano para el Desarrollo de la Ciencia y la Tecnología “Francisco José de Caldas” COLCIENCIAS-CENIVAM (Contract No. 432-2004). The authors acknowledge also the partial financial support from Spanish “*Comisión Interministerial de Ciencia y Tecnología*” (CICYT) (Project Reference: SAF2006-04698). Finally, but not least, this work was supported in part by VLIR (Vlaamse InterUniversitaire Raad, Flemish Interuniversity Council, Belgium) under the IUC Program VLIR-UCLV.

## 5. Reference and Notes

1996. The Merck Index. Chapman and Hall.
- Aguirre, G., Boiani, M., Cerecetto, H., Gerpe, A., González, M., Fernández-Sainz, Y., Denicola, A., Ochoa, C., Nogal, J.J., Montero, D., Escario, J.A., 2004. Novel Antiprotozoal Products: Imidazole and Benzimidazole N-Oxide Derivatives and Related Compounds. *Arch. Pharm.* 337, 259-270.
- Andrade, S.G., Rassi, A., Magalhaes, J.B., Filho, F.F., Luquetti, A.O., 1992. Specific chemotherapy of Chagas disease: a comparison between the response in patients and experimental animals inoculated with the same strains. *Trans. R. Soc. Tropical Med. Hygiene* 86, 624-626.
- Bal, G., Van der Veken, P., Antonov, D., Lambeir, A.-M., Grellier, P., Croft, S.L., Augustyns, K., Haemers, A., 2003. Prolylisoaxazoles: potent inhibitors of prolyl oligopeptidase with antitrypanosomal activity. *Bioorg. Med. Chem. Lett.* 13, 2875-2878.
- Baldi, P., Brunak, S., Chauvin, Y., Andersen, C.A., Nielsen, H., 2000. Assessing the accuracy of prediction algorithms for classification: an overview. *Bioinformatics* 16, 412-424.
- Bonse, S., Santelli-Rouvier, C., Barbe, J., Krauth-Siegel, R.L., 1999. Inhibition of *Trypanosoma cruzi* Trypanothione Reductase by Acridines: Kinetic Studies and Structure-Activity Relationships. *J. Med. Chem.* 42, 5448-5454.
- Buckner, F., Yokoyama, K., Lockman, J., Aikenhead, K., Ohkanda, J., Sadilek, M., Sebti, S.d., Van Voorhis, W., Hamilton, A., Gelb, M.H., 2003. A class of sterol 14-demethylase inhibitors as anti-*Trypanosoma cruzi* agents. *Proc. Nat. Acad. Sci. USA* 100, 15149-15153.
- Buckner, F.S., Griffin, J.H., Wilson, A.J., Van Voorhis, W.C., 2001. Potent Anti-*Trypanosoma cruzi* Activities of Oxidosqualene Cyclase Inhibitors. *Antimicrob. Agents Chemother.* 45, 1210-1215.

- Casañola-Martin, G.M., Khan, M.T.H., Marrero-Ponce, Y., Ather, A., Sultankhodzhaev, M.N., Torrens, F., 2006. New Tyrosinase Inhibitors Selected by Atomic Linear Indices-Based Classification Models. *Bioorg. Med. Chem. Lett.* 16, 324-330.
- Casanola-Martin, G.M., Marrero-Ponce, Y., Khan, M.T., Ather, A., Sultan, S., Torrens, F., Rotondo, R., 2007. TOMOCOMD-CARDD descriptors-based virtual screening of tyrosinase inhibitors: evaluation of different classification model combinations using bond-based linear indices. *Bioorg Med Chem* 15, 1483-1503.
- Castillo-Garit, J.A., Marrero-Ponce, Y., Escobar, J., Torrens, F., Rotondo, R., 2008a. A novel approach to predict aquatic toxicity from molecular structure. *Chemosphere* 73, 415-427.
- Castillo-Garit, J.A., Marrero-Ponce, Y., Torrens, F., 2006. Atom-based 3D-chiral quadratic indices. Part 2: prediction of the corticosteroid-binding globulinbinding affinity of the 31 benchmark steroids data set. *Bioorg. Med. Chem.* 14, 2398-2408.
- Castillo-Garit, J.A., Marrero-Ponce, Y., Torrens, F., García-Domenech, R., 2008b. Estimation of ADME Properties in Drug Discovery: Predicting Caco-2 Cell Permeability Using Atom-Based Stochastic and Non-Stochastic Linear Indices. *J. Pharm. Sci.* 97, 1946-1976.
- Castillo-Garit, J.A., Marrero-Ponce, Y., Torrens, F., García-Domenech, R., Romero-Zaldivar, V., 2008c. Bond-Based 3D-Chiral Linear Indices: Theory and QSAR Applications to Central Chirality Codification. *J. Comput. Chem.* 29, 2500 - 2512.
- Castillo-Garit, J.A., Marrero-Ponce, Y., Torrens, F., Rotondo, R., 2007. Atom-based Stochastic and non-Stochastic 3D-Chiral Bilinear Indices and their Applications to Central Chirality Codification. *J. Mol. Graphics Model.* 26, 32-47.
- Castillo Garit, J.A., Martinez-Santiago, O., Marrero Ponce, Y., Casañola-Martin, G.M., Torrens, F., 2008. Atom-based non-stochastic and stochastic bilinear indices: Application to QSPR/QSAR studies of organic compounds. *Chem. Phys. Lett.* 464, 107-112.
- Cerectto, 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., Basombrío, M.A., 2000. Synthesis and antitrypanosomal evaluation of E-isomers of 5-nitro-2-furaldehyde and 5-nitrothiophene-2-carboxaldehyde semicarbazone derivatives. Structure-activity relationships. *Eur. J. Med. Chem.* 35, 343-350.
- Consonni, V., Todeschini, R., Pavan, M., 2002. Structure/Response Correlations and Similarity/Diversity Analysis by GETAWAY Descriptors. 1. Theory of the Novel 3D Molecular Descriptors. *J. Chem. Inf. Comput. Sci.* 42, 682-692.
- COSTB9, Action on Chemotherapy of Protozoal Infections 1997–2002. Final report. European Cooperation in the Field of Scientific and Technical Research (COST). European Commission.
- Daunes, S., D'Silva, C., 2002. The therapeutic potential of inhibitors of the trypanothione cycle. *Expert Opin. Invest. Drugs* 11, 217-231.
- Du, X., Guo, C., Hansell, E., Doyle, P.S., Caffrey, C.R., Holler, T.P., McKerrow, J.H., Cohen, F.E., 2002. Synthesis and Structure-Activity Relationship Study of Potent Trypanocidal Thio Semicarbazone Inhibitors of the Trypanosomal Cysteine Protease Cruzain. *J. Med. Chem.* 45, 2695-2707.
- Elhalem, E., Bailey, B.N., Docampo, R., Ujvary, I., Szajnman, S.H., Rodriguez, J.B., 2002. Design, Synthesis, and Biological Evaluation of Aryloxyethyl Thiocyanate Derivatives against *Trypanosoma cruzi*. *J. Med. Chem.* 45, 3984-3999.
- Estrada, E., 1999. In: Devillers, J., Balaban, A.T. (Eds.). *Topological Indices and Related Descriptors in QSAR and QSPR*. Gordon and Breach, Amsterdam, pp. 403–453.

- Estrada, E., Peña, A., 2000. In Silico Studies for the Rational Discovery of Anticonvulsant Compounds. *Bioorg. Med. Chem.* 8, 2755-2770.
- Estrada, E., Uriarte, E., Montero, A., Teijeira, M., Santana, L., De Clercq, E., 2000. A Novel Approach for the Virtual Screening and Rational Design of Anticancer Compounds. *J. Med. Chem.* 43, 1975-1985.
- Faundez, M., Pino, L., Letelier, P., Ortiz, C., Lopez, R., Seguel, C., Ferreira, J., Pavani, M., Morello, A., Maya, J.D., 2005. Buthionine Sulfoximine Increases the Toxicity of Nifurtimox and Benznidazole to *Trypanosoma cruzi*. *Antimicrob. Agents Chemother.* 49, 126-130.
- Gilbert, I.H., 2002. Inhibitors of dihydrofolate reductase in leishmania and trypanosomes. *Biochim. Biophys. Acta* 1587, 249-257.
- Gillmor, S.A., Craik, C.S., Fletterick, R.J., 1997. Structural determinants of specificity in the cysteine protease cruzain. *Protein Sci.* 6, 1603-1611.
- Gómez-Barrio, A., Martínez-Díaz, R.A., Atienza, J., Escario, J.A., Diego, C., Avendaño, C., 1997. New derivatives of gentian violet as trypanocides: In vitro and in vivo assays on *Trypanosoma cruzi*. *Res. Rev. Parasitol.* 57, 25-31.
- Hamilton, C.J., Saravanamuthu, A., Eggleston, I.M., Fairlamb, A.H., 2003. Ellman's-reagent-mediated regeneration of trypanothione in situ: substrate-economical microplate and time-dependent inhibition assays for trypanothione reductase. *Biochem. J.* 369, 529-537.
- Hiyoshi, H., Yanagimachi, M., Ito, M., Ohtsuka, I., Yoshida, I., Saeki, T., Tanaka, H., 2000. Effect of ER-27856, a novel squalene synthase inhibitor, on plasma cholesterol in rhesus monkeys: comparison with 3-hydroxy-3-methylglutaryl-CoA reductase inhibitors. *J. Lipid Res.* 41, 1136-1144.
- Huang, L., Brinen, L.S., Ellman, J.A., 2003. Crystal structures of reversible ketone-Based inhibitors of the cysteine protease cruzain. *Bioorg. Med. Chem.* 11, 21-29.
- Kaminsky, R., Brun, R., 1993. In vitro assays to determine drug sensitivities of African trypanosomes: a review. *Acta Trop.* 54, 279-289.
- Kier, L.B., Hall, L.H., 1986. *Molecular Connectivity in Structure-Activity Analysis*. Research Studies Press, Letchworth, U. K.
- Marrero-Ponce, Y., 2003. Total and Local Quadratic Indices of the Molecular Pseudograph's Atom Adjacency Matrix: Applications to the Prediction of Physical Properties of Organic Compounds. *Molecules* 8, 687-726.
- Marrero-Ponce, Y., 2004. Total and Local (Atom and Atom-Type) Molecular Quadratic Indices: Significance-Interpretation, Comparison to Other Molecular Descriptors, and QSPR/QSAR Applications. *Bioorg. Med. Chem.* 12, 6351-6369.
- Marrero-Ponce, Y., Cabrera, M.A., Romero-Zaldivar, V., Bermejo, M., Siverio, D., Torrens, F., 2005a. Prediction of Intestinal Epithelial Transport of Drug in (Caco-2) Cell Culture from Molecular Structure using in silico Approaches During Early Drug Discovery. *Internet Electron. J. Mol. Des.* 4 124-150.
- Marrero-Ponce, Y., Castillo-Garit, J.A., 2005. 3D-chiral Atom, Atom-type, and Total Non-Stochastic and Stochastic Molecular Linear Indices and Their Applications to Central Chirality Codification. *J. Comput. Aided Mol. Des.* 19, 369-383.
- Marrero-Ponce, Y., Castillo-Garit, J.A., Olazabal, E., Serrano, H.S., Morales, A., Castanedo, N., Ibarra-Velarde, F., Huesca-Guillen, A., Sanchez, A.M., Torrens, F., Castro, E.A., 2005b. Atom, Atom-Type and Total Molecular Linear Indices as a Promising Approach for Bioorganic and Medicinal Chemistry: Theoretical and Experimental Assessment of a Novel Method for Virtual Screening and Rational Design of New Lead Anthelmintic. *Bioorg. Med. Chem.* 13, 1005-1020.

- Marrero-Ponce, Y., Castillo-Garit, J.A., Torrens, F., Romero-Zaldivar, V., Castro, E., 2004. Atom, Atom-Type, and Total Linear Indices of the "Molecular Pseudograph's Atom Adjacency Matrix": Application to QSPR/QSAR Studies of Organic Compounds. *Molecules* 9, 1100-1123.
- Marrero-Ponce, Y., Huesca-Guillen, A., Ibarra-Velarde, F., 2005c. Quadratic Indices of the "Molecular Pseudograph's Atom Adjacency Matrix" and Their Stochastic Forms: a Novel Approach for Virtual Screening and in silico Discovery of New Lead Paramphistomicide Drugs-like Compounds. *J. Mol. Struct. (Theochem)* 717, 67-79.
- Marrero-Ponce, Y., Iyarreta-Veitia, M., Montero-Torres, A., Romero-Zaldivar, C., Brandt, C.A., Avila, P.E., Kirchgatter, K., Machado, Y., 2005d. Ligand-Based Virtual Screening and in silico Design of New Antimalarial Compounds Using Nonstochastic and Stochastic Total and Atom-Type Quadratic Maps. *J. Chem. Inf. Comput. Sci.* 45, 1082-1100.
- Marrero-Ponce, Y., Khan, M.T.H., Casañola-Martín, G.M., Ather, A., Sultankhodzhaev, M.N., Torrens, F., Rotondo, R., 2007. Prediction of Tyrosinase Inhibition Spectra for Chemicals Using Novel Atom-Based Bilinear Indices. *ChemMedChem* 2, 449 – 478.
- Marrero-Ponce, Y., Marrero, R.M., Torrens, F., Martinez, Y., Bernal, M.G., Zaldivar, V.R., Castro, E.A., Abalo, R.G., 2006a. Non-stochastic and stochastic linear indices of the molecular pseudograph's atom-adjacency matrix: a novel approach for computational in silico screening and "rational" selection of new lead antibacterial agents. *J. Mol. Model.* 12, 255-271.
- Marrero-Ponce, Y., Medina-Marrero, R., Castillo-Garit, J.A., Romero-Zaldivar, V., Torrens, F., Castro, E.A., 2005e. Protein Linear Indices of the "Macromolecular Pseudograph Alpha-Carbon Atom Adjacency Matrix" in Bioinformatics. Part 1: Prediction of Protein Stability Effects of a Complete Set of Alanine Substitutions in Arc Repressor. *Bioorg. Med. Chem.* 13, 3003-3015.
- Marrero-Ponce, Y., Meneses-Marcel, A., Catillo-Garit, J.A., Machado-Tugores, Y., Escario, J.A., Gómez-Barrio, A., Montero Pereira, D., Nogal-Ruiz, J.J., Arán, V.J., Martínez-Fernández, A.R., Torrens, F., Rotondo, R., 2006b. Predicting Antitrichomonal Activity: A Computational Screening Using Atom-Based Bilinear Indices and Experimental Proofs. *Bioorg. Med. Chem.* 14, 6502–6524.
- Marrero-Ponce, Y., Romero, V., 2002. TOMOCOMD-CARDD software. TOMOCOMD (TOpological MOlecular COMputer Design) for Windows, version 1.0 is a preliminary experimental version; in future a professional version can be obtained upon request to Y. Marrero: yovanimp@uclv.edu.cu or ymarrero77@yahoo.es Central University of Las Villas, Santa Clara, Villa Clara.
- Marrero-Ponce, Y., Torrens, F., Alvarado, Y.J., Rotondo, R., 2006c. Bond-Based Global and Local (Bond, Group and Bond-Type) Quadratic Indices and Their Applications to Computer-Aided Molecular Design. 1. QSPR Studies of Diverse Sets of Organic Chemicals. *J. Comput. Aided Mol. Des.* 20, 685–701.
- Marrero-Ponce, Y., Torrens, F., García-Domenech, R., Ortega-Broche, S.E., Romero Zaldivar, V., 2008. Novel 2D TOMOCOMD-CARDD Descriptors: Atom-Based Stochastic and Non-Stochastic Bilinear Indices and Their QSPR Applications. *J. Math. Chem.* 44, 650-673.
- Marrero Ponce, Y., 2004. Linear Indices of the "Molecular Pseudograph's Atom Adjacency Matrix": Definition Significance-Interpretation, and Application to QSAR Analysis to Flavone Derivatives as HIV-1 Integrase Inhibitors. *J. Chem. Inf. Comput. Sci.* 44, 2010-2026.
- Marrero Ponce, Y., Castillo Garit, J.A., Nodarse, D., 2005. Linear Indices of the "Macromolecular Graph's Nucleotides Adjacency Matrix' as a Promising Approach for Bioinformatics Studies.



- Part 1: Prediction of Paromomycin's Affinity Constant with HIV-1  $\psi$ -RNA Packaging Region. *Bioorg. Med. Chem.* 13, 3397-3404.
- Negwer, M., 1987. *Organic-Chemical Drugs and their Synonyms*. Akademie, Berlin.
- Pauling, L., 1939. *The Nature of Chemical Bond*. Cornell University Press Ithaca (NY).
- Prieto, J., Talevi, A., Bruno-Blanch, L., 2006. Application of linear discriminant analysis in the virtual screening of antichagasic drugs through trypanothione reductase inhibition. *Mol. Div.* 10, 361-375.
- Renslo, A.R., McKerrow, J.H., 2006. Drug Discovery and Development for Neglected Parasitic Diseases. *Nat. Chem. Biol.* 2, 701-710.
- Rolón, M., Seco, E.M., Vega, C., Nogal, J.J., Escario, J.A., Gómez-Barrio, A., Malpartida, F., 2006a. Selective activity of polyene macrolides produced by genetically modified *Streptomyces* on *Trypanosoma cruzi*. *Int. J. Antimicrob. Agents* 28, 104-109.
- Rolón, M., Vega, C., Escario, J., Gómez-Barrio, A., 2006b. Development of resazurin microtiter assay for drug sensibility testing of *Trypanosoma cruzi* epimastigotes. *Parasitol. Res.* 99, 103-107.
- Salmon-Chemin, L., Buisine, E., Yardley, V., Kohler, S., Debreu, M.A., Landry, V., Sergheraert, C., Croft, S.L., Krauth-Siegel, R.L., Davioud-Charvet, E., 2001. 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, 548-565.
- Silva, L.H., Nussensweig, V., 1953. Sobre uma cepa *Trypanosoma cruzi* virulenta para o camundongo branco. *Folia of Clinical Biology* 20, 191-207.
- STATISTICA version. 6.0, 2001. StatSoft, Tulsa.
- Todeschini, R., Gramatica, P., 1998. New 3D Molecular Descriptors: the WHIM Theory and QSAR Applications. *Perspect. Drug Disc.* 9-11, 355-380.
- Urbina, J.A., 2002. Chemotherapy of Chagas Disease. *Current Pharm. Design* 8, 287-295.
- Urbina, J.A., Concepcion, J.L., Caldera, A., Payares, G., Sanoja, C., Otomo, T., Hiyoshi, H., 2004. In Vitro and In Vivo Activities of E5700 and ER-119884, Two Novel Orally Active Squalene Synthase Inhibitors, against *Trypanosoma cruzi*. *Antimicrob. Agents Chemother.* 48, 2379-2387.
- Urbina, J.A., Docampo, R., 2003. Specific chemotherapy of Chagas disease: controversies and advances. *Trends Parasitol.* 19, 495-501.
- Urbina, J.A., Payares, G., Contreras, L.M., Liendo, A., Sanoja, C., Molina, J., Piras, M., Piras, R., Perez, N., Wincker, P., Loebenberg, D., 1998. Antiproliferative Effects and Mechanism of Action of SCH 56592 against *Trypanosoma* (*Schizotrypanum*) *cruzi*: In Vitro and In Vivo Studies. *Antimicrob. Agents Chemother.* 42, 1771-1777.
- Vega, M.C., Rolon, M., Martinez-Fernandez, A.R., Escario, J.A., Gomez-Barrio, A., 2005. A New Pharmacological Screening Assay with *Trypanosoma cruzi* Epimastigotes Expressing Beta-Galactosidase. *Parasitol. Res.* 95, 296-298.
- Weisman, J.L., Liou, A.P., Shelat, A.A., Cohen, F.E., Guy, R.K., DeRisi, J.L., 2006. Searching for New Antimalarial Therapeutics Amongst Known Drugs. *Chem. Biol. Drug Des.* 67, 409-416.
- Werbovetz, K.A., 2000. Target-Based Drug Discovery for Malaria, Leishmaniasis, and Trypanosomiasis. *Current Med. Chem.* 7, 835-860.
- Zuccotto, F., Zvelebil, M., Brun, R., Chowdhury, S.F., Di Lucrezia, R., Leal, I., Maes, L., Ruiz-Perez, L.M., Gonzalez Pacanowska, D., Gilbert, I.H., 2001. Novel inhibitors of *Trypanosoma cruzi* dihydrofolate reductase. *Eur. J. Med. Chem.* 36, 395-405.

## Supporting Information

# Computational Discovery of Novel Tripanosomicidals Drug-like Chemicals by Using Bond-based Non-stochastic and Stochastic Quadratic Maps and Linear Discriminant Analysis.

Juan Alberto Castillo-Garit,<sup>a,b,c\*</sup> Maria C. Vega,<sup>d</sup> Miriam Rolon,<sup>d</sup> Yovani Marrero-Ponce,<sup>b,c,f</sup> Vladimir V. Kouznetsov,<sup>g</sup> Diego Fernando Amado Torres,<sup>g</sup> Alicia Gómez-Barrio,<sup>d</sup> Alfredo Alvarez Bello,<sup>b</sup> Alina Montero,<sup>b</sup> Francisco Torrens,<sup>c</sup> and Facundo Pérez-Giménez.<sup>f</sup>

### Table of Contents

---

<b>Table S1.</b> Classification of active compounds included in the training set using Equations <b>1</b> and <b>2</b> .....	<b>2</b>
<b>Table S2.</b> Classification of inactive compounds included in the training set using Equations <b>1</b> and <b>2</b> .....	<b>4</b>
<b>Table S3.</b> Classification of active compounds included in the test set using Equations <b>1</b> and <b>2</b> .....	<b>7</b>
<b>Table S4.</b> Classification of inactive compounds included in the test set using Equations <b>1</b> and <b>2</b> .....	<b>8</b>
<b>Figure S1</b> Molecular structures of anti-trypanosomal compounds used in training set.....	<b>9</b>
<b>Figure S2.</b> Molecular structures of anti-trypanosomal compounds used in test set.....	<b>14</b>

---

**Table S1.** Classification of active compounds included in the training set using Equations 1 and 2

Compounds	Eq.1	Eq. 2
1e	63.0	58.3
2e	86.5	37.1
3e	94.7	71.6
4e	98.1	86.2
5e	33.6	8.6
6e	68.5	41.0
7e	96.5	60.9
8e	81.0	90.1
9e	99.9	57.3
10e	69.0	97.1
11e	98.3	95.7
12e	89.6	91.3
13e	93.8	-93.7
14e	98.6	38.5
15e	94.5	-78.0
16e	99.9	98.7
17e	99.9	99.2
18e	100.0	98.6
19e	99.9	99.6
20e	99.9	99.5
21e	99.0	98.7
22e	94.2	76.3
23e	92.2	28.7
24e	94.1	71.4
25e	93.9	99.9
25e	100.0	96.0
27e	99.7	94.0
28e	99.7	92.4
29e	100.0	95.4
30e	74.9	90.3
31e	63.2	98.8
32e	64.8	43.1
33e	59.9	54.1
34e	82.9	78.3

**Table S1.** *Cont...*

Compounds	Eq.1	Eq. 2
35e	80.3	84.2
36e	99.8	60.1
37e	99.7	66.0
38e	99.9	99.0
39e	99.9	98.8
40e	98.3	93.4
41e	97.8	69.1
42e	99.4	94.8
43e	100.0	100.0
44e	25.5	97.9
45e	74.7	99.8
46e	-47.30	91.34
47e	95.32	96.91
48e	84.93	22.81
49e	97.39	99.39
50e	99.19	98.92
51e	63.86	-15.01
52e	61.84	-28.59
53e	65.14	51.68
54e	44.20	91.72
55e	36.70	90.57
56e	43.34	79.92
57e	48.37	89.63
58e	93.21	95.20
59e	92.04	96.10
60e	92.11	95.80
61e	87.82	96.92
62e	89.76	99.40
63e	90.75	97.11
64e	95.10	67.74
65e	94.70	87.58
66e	99.92	99.90
67e	99.98	99.97
68e	99.99	96.23
69e	99.96	98.84
70e	98.51	92.53
71e	99.95	83.14
72e	99.96	99.99
73e	90.32	95.67

**Table S1.** *Cont...*

Compounds	Eq.1	Eq. 2
74e	81.94	10.84
75e	75.92	57.03
76e	65.11	38.57
77e	87.17	75.88
78e	56.46	-23.08
79e	63.18	-6.57
80e	99.48	97.76
81e	99.52	98.36
82e	74.50	-70.67
83e	99.98	98.39
84e	99.41	96.57
85e	82.11	55.45
86e	90.07	60.23
87e	93.93	94.20
88e	20.63	95.03
89e	83.05	-6.58
90e	89.27	38.49
91e	99.64	40.55
92e	99.62	97.71
93e	99.88	11.64
94e	99.31	13.34
95e	99.21	8.28
96e	99.39	7.96
97e	94.52	18.44
98e	95.54	82.64
99e	82.26	74.49
100e	91.85	73.05
101e	97.65	82.40
102e	-26.67	-13.89
103e	3.98	-82.54
104e	42.86	-23.50
105e	69.34	-16.31
106e	44.20	91.72
107e	98.84	26.69
108e	99.72	71.02
109e	99.36	20.89
110e	99.43	92.52
111e	88.20	30.51
112e	91.31	-18.74

**Table S1.** *Cont...*

Compounds	Eq.1	Eq. 2
113e	85.19	-78.34
114e	98.89	99.99
115e	85.10	82.60
116e	95.64	33.95
117e	88.99	71.46
118e	87.63	81.78
119e	73.90	-19.75
120e	77.96	99.68

**Table S2** Classification of inactive compounds included in the training set using Equations 1 and 2

Compounds	Eq.1	Eq. 2
Va1-11	82.59	-30.71
Va1-13	92.32	-52.40
Va1-15	98.15	55.10
Va1-31	61.82	-50.77
Va1-33	83.22	-44.74
Va2-10	89.80	80.25
Va2-17	97.84	75.34
Va2-26	26.65	-49.74
Va2-27	96.83	64.56
Va3-3c	-80.66	-44.06
Va3-3f	-6.87	-22.08
Va5-9a	96.29	45.00
Va5-10	96.55	25.86
Va5-14j	99.31	59.25
Va5-14g	99.75	92.25
Va6-5b	93.67	-35.17
Va6-5d	98.48	74.84
Va6-9a	83.99	-21.23
Va6-10a	81.91	6.68
Vax-6	90.65	-52.76
Vax-7	92.06	-52.15
Vax-13	-49.53	-61.64
Amantadine	-96.14	-51.27
3-Episiostatin B	-99.42	-82.48
PALA	-99.82	-81.26
Thiacetazone	-94.09	-3.14
TBHQ	-94.09	-40.66
BL-1743	-95.55	-70.85
Dimepranol	-98.46	-95.73
LK-274	-98.68	-89.51
Ethoxene	-91.88	-88.87
Ketoxal	-99.21	-96.18
DANA	-99.51	-52.68
Moroxidine	-99.24	-94.77
Brometenamine	-98.98	-62.67
Cloral betaine	-99.89	-81.21
Triclofos	-99.95	-86.00
Trichlorourethan	-99.78	-99.54
Isopral	-99.59	-88.65

**Table S2.** *Cont...*

Compounds	Eq.1	Eq. 2
Thiourethane	-97.45	-96.02
Urethane	-97.86	-96.45
Etiron	-96.93	-97.15
Bromobutanol	-99.88	-99.89
Aleudrin"Beckmann"	-99.91	-78.75
Chlorobutanol	-99.92	-95.17
Vernelan	-98.73	-89.03
Agr 449	-97.03	-90.85
Sedanfactor		
Solucion	-95.79	-91.62
Cloretate	-99.97	-95.95
Alcabrol	-98.58	-39.46
Valerium Paul		
Thibault	-93.01	-94.50
Methonal	-99.66	-95.34
Nitroinosite	-100.0	67.09
Cetohexazine	-90.19	-83.43
Bason	-98.98	-84.76
Methylpentynol	-94.66	-90.33
Bromisoval	-99.39	-77.05
Carbromide	-98.79	-92.82
Ethchlorvynol	-98.51	-81.43
Pentrichloral	-99.85	-81.14
Mepentamate	-96.87	-80.58
Carbavin	-97.46	-83.07
Phenacemide	-89.73	-24.16
Atrolactamide	-95.92	-46.25
Tetharbital	-99.15	-74.91
Brofoxine	-97.80	96.38
Norantoin	-91.17	-54.19
Sultiame	-94.99	-53.48
Albutoin	-93.19	-95.05
Athotoin	-87.64	-56.44
Nirvanol	-92.16	-28.50
Fenaclon	-94.28	-37.79
Pheneturide	-88.86	7.09
Tetrantoin	-86.73	-0.11
Phenobarbital	-94.68	5.26
Mesuximide	-90.99	0.35

**Table S2. Cont...**

Compounds	Eq.1	Eq. 2
Phenethylone	-89.12	-11.37
Metetoin	-94.44	-32.47
Mephentyoin	-94.02	-24.22
Orotosan Fe	-98.74	-84.69
Cobalti glutamas	-98.43	-86.68
Ferrocholate	-99.50	-95.90
Ferrosi ascorbas	-99.37	-78.57
Arecoline	-97.06	-82.09
Butanolum	-92.08	-94.81
Cupriaseptol	-95.49	-86.81
Etamsylate	-98.06	-81.79
Sango-Stop	-99.57	-82.66
Carbazochrome sodium sulfonate	-99.04	74.06
A-Peest	-97.99	29.37
Besunide	-79.94	-9.82
Vitamin B15	-99.85	-94.69
Spirolactone	-99.66	98.86
Chlormetaminofenamide	-99.80	-73.41
Lemidosul	-98.20	-71.08
Urea	-98.03	-94.78
Guanamine	-98.35	-94.27
Glycerol	-98.33	-89.54
Succinic acid	-97.73	-88.99
Pamabron	-98.66	-94.95
Oxaden	-98.56	-88.83
Propazolamide	-99.58	-95.78
Methazolamide	-99.75	-96.25
Chlormerodrin	-99.91	-75.18
Theophylline-Merodrin	-99.67	-89.23
Clofenamide	-99.71	-72.19
Chloraminophenamidum	-99.83	-76.56
Butazolamidum	-99.50	-97.88
Ambuside	-99.61	-26.83
Merbiurelidin	-99.88	-90.67
Manna sugar	-99.65	-84.72
Chlorothiazide	-99.67	-69.32

**Table S2. Cont...**

Compounds	Eq.1	Eq. 2
Iodothiazide	-99.45	-11.93
Carzenide	-96.10	-53.25
Teofillina	-99.37	-89.17
Siccamid	-99.80	-59.73
Teclotiazide	-99.70	-76.67
Trichlormethiazide	-99.95	80.28
Propamin"soviet	-94.22	-90.86
Cystamine	-97.89	-74.17
Amifostine	-99.67	-99.51
Adeturon	100.00	-72.10
Batilol	-87.08	-72.56
Glisolamide	-93.82	-39.45
Glibutimine	-71.61	-43.86
RMI 11894	-74.13	-62.41
Tonyol	-99.63	-86.72
Tiforminhydrochloride	-98.27	-91.14
Etoformin hydrochloride	-98.47	-99.09
Benfosformin	-99.60	-44.01
Phenformin hydrochloride	-92.81	-52.40
Mebenformin	-95.85	-56.03
Glycopyramide	-98.82	-65.52
Glyprothiazol	-94.10	-85.19
Phenbutamide	-93.64	-87.57
Carbutamide	-95.36	-93.53
Metasulfanilbutylcarbamide	-95.49	-92.67
Centpiperalone	-88.26	-38.81
Glybuzole	-96.72	-52.74
SPC-703	-91.90	-84.93
Tolpyramide	-89.72	-87.67
Glybuthiazol	-97.62	-72.97
Thiazanol	-97.21	-94.00
Tolbutamide	-34.79	-91.67
Hypoglycin B	-98.29	-32.85
Anticomane	-97.97	-93.08
Furfurvlurea	-39.71	-33.51



Table S2. Cont...

Compounds	Eq.1	Eq. 2
Glymidine sodium	-98.18	-3.55
Clorexamide	-97.31	-63.72
Glysobuzole	-96.26	-92.79
Glypinamide	-98.54	-78.21
Ag 307	-98.51	-78.83
Bromcholine	-99.78	-98.61
Meбетide	-90.12	-70.26
Minoxidil	-97.27	-91.71
Hemedin	-96.17	-92.97
Penbutamin	-99.85	-95.12
Guanoctine hydrochloride	-99.43	-91.11
Guanisoquine sulfate	-94.16	96.30
Flutonidine hydrochloride	-98.51	-37.72
Nicopholine	-92.73	-65.23
Gangliefene hydrochloride	-82.64	-49.33
Quateron	-77.36	-52.22
Pancuronium bromide	-97.78	87.27
Plegarol	-95.59	-92.71
Gaplegin	-99.43	-96.62
Oxaditon	-99.56	-96.23
Agentit	-99.74	-97.02
Dimecolonium iodide	-99.66	-91.80
Dicoline	-97.23	-92.12
Methyloxamethonium iodide	-99.68	-96.95
Hexamethonium	-99.86	-96.32
Trepirium iodide	-99.68	-91.63
Tiamethonium iodide	-99.73	-97.72
Imekhin	-98.49	-83.85
Dimecamine	-99.86	-68.44
Penhexamine	-99.67	-87.39
Leptodactyline	-96.80	-33.07

Table S2. Cont...

Compounds	Eq.1	Eq. 2
Tetrammonii iodidum	-99.63	-95.35
Cyclocholine tosilate	-92.72	-94.95
Methylene chloride	-99.61	-93.26
Ethylene	-88.27	-90.73
Dioxychlorane	-99.97	-75.68
Aliflurane	-99.90	-82.69
Vinyl ether	-93.33	-91.45
Tiouracilo	-96.41	-86.31
Thiamazole	-97.28	-90.74
Carbimazole	-98.88	-96.45
Perchloroetane	-99.98	-64.69
Lindane	-99.91	-31.05
Ascaridole	-97.27	-83.09
Pyrantel tartrate	-84.94	-69.72
Fentanilo	58.49	76.89
Tenalidine tartrate	-67.31	26.02
Dioxoprometazine hydrochloride	-83.61	63.75
N-hidroxymetil-N-metilurea	-99.20	-93.33
2,4,5 triclofenol (2-hidroxypropyl trimetilamonium hydroxide	-99.76	-6.05
Norgamem	-99.42	-97.35
Emylcamate	-98.17	-93.21
Mephenesin carbamate	-98.12	-89.89
Methocarbamol	-95.72	-50.83
Luvatren	-98.71	-58.94
Lorbamate	-88.51	28.83
Furtrethonium iodide	-99.04	-37.69
Ergocristine	-98.75	-93.08
Etilefrine divalate	7.27	97.78
	-95.38	-25.78

**Table S2.** *Cont...*

Compounds	Eq.1	Eq. 2
Etafedrine		
hydrochloride	-88.38	-50.23
Isoprenaline	-94.55	-44.83
Isofenefrine	-88.63	-47.75
Phenylethanolamine		
Adrenaline borate	-96.35	-46.32
Amfetamine	-61.90	-56.61
Norephedrine	-82.82	-55.21
Metaraminol	-90.25	-57.43
Synephrine	-91.90	-63.53
Corbadrine	-95.35	-54.42
Nibroxane	-99.86	-79.67
Alafosfalin	-99.84	-97.41
Furaguanidine	-99.65	-64.96
Sulfanilamide	-94.07	-85.82
Parabortine	-97.11	29.68
Allicin	-92.65	-84.77
1442 Streptomycin	-99.88	22.37
1443		
Azirinomycin	-95.54	-88.18
1444 Gentamicin		
A	-99.32	-34.84

**Table S3.** Classification of active compounds included in the test set using Equations 1 and 2

Compounds	Eq.1	Eq. 2
1p	84.32	78.61
2p	92.80	82.28
3p	88.31	65.91
4p	99.17	88.68
5p	97.23	41.36
6p	99.93	98.03
7p	99.94	97.88
8p	99.92	99.33
9p	99.94	98.08
10p	99.90	99.73
11p	55.44	99.15
12p	78.78	90.86
13p	58.27	99.43
14p	87.94	-31.60
15p	86.34	-45.76
16p	83.05	73.11
17p	34.32	74.36
18p	98.74	95.50
19p	55.58	29.18
20p	85.33	61.55
21p	98.05	89.12
22p	95.66	37.25
23p	99.23	64.85

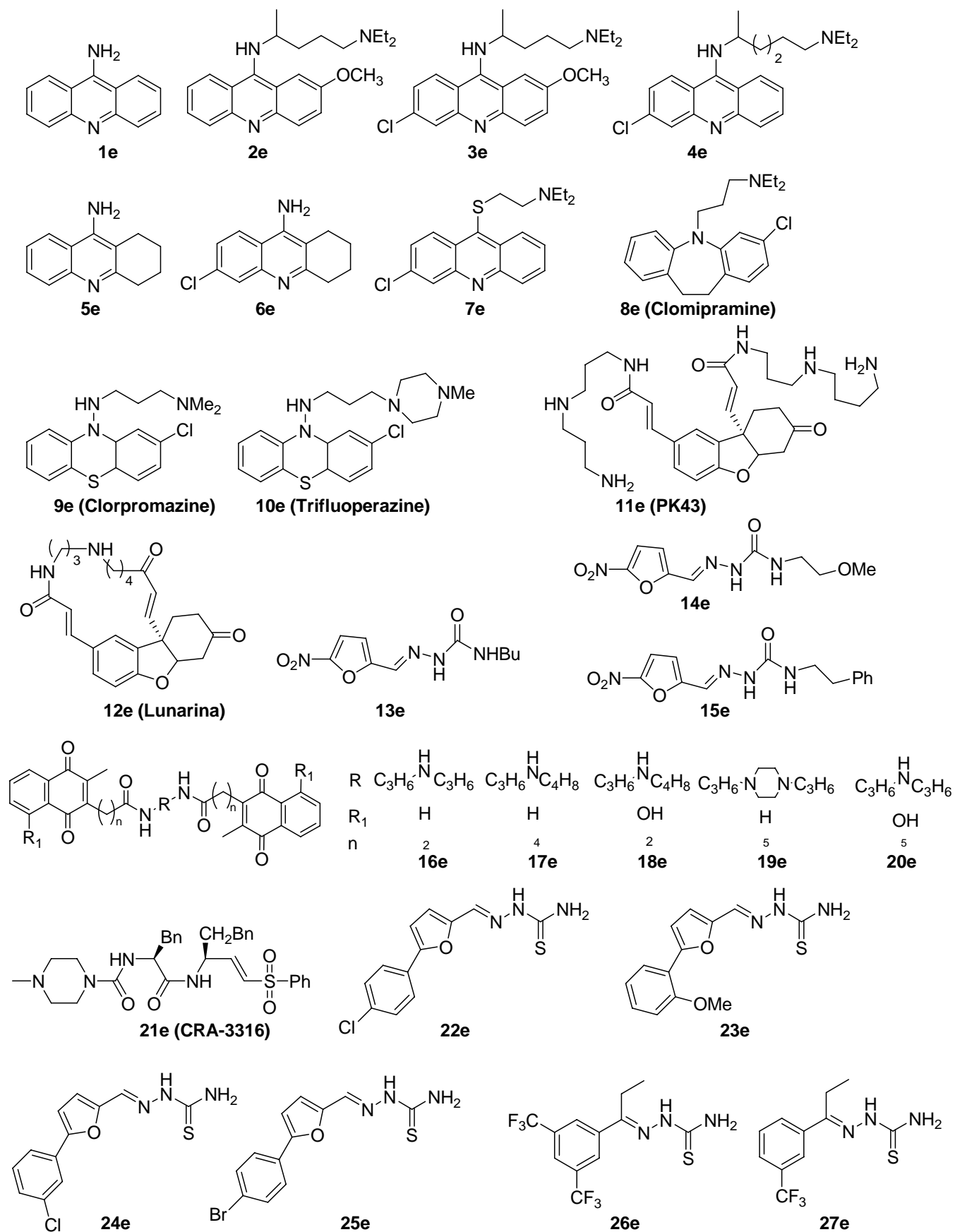


**Table S4.** Classification of inactive compounds included in the test set using Equations 1 and 2

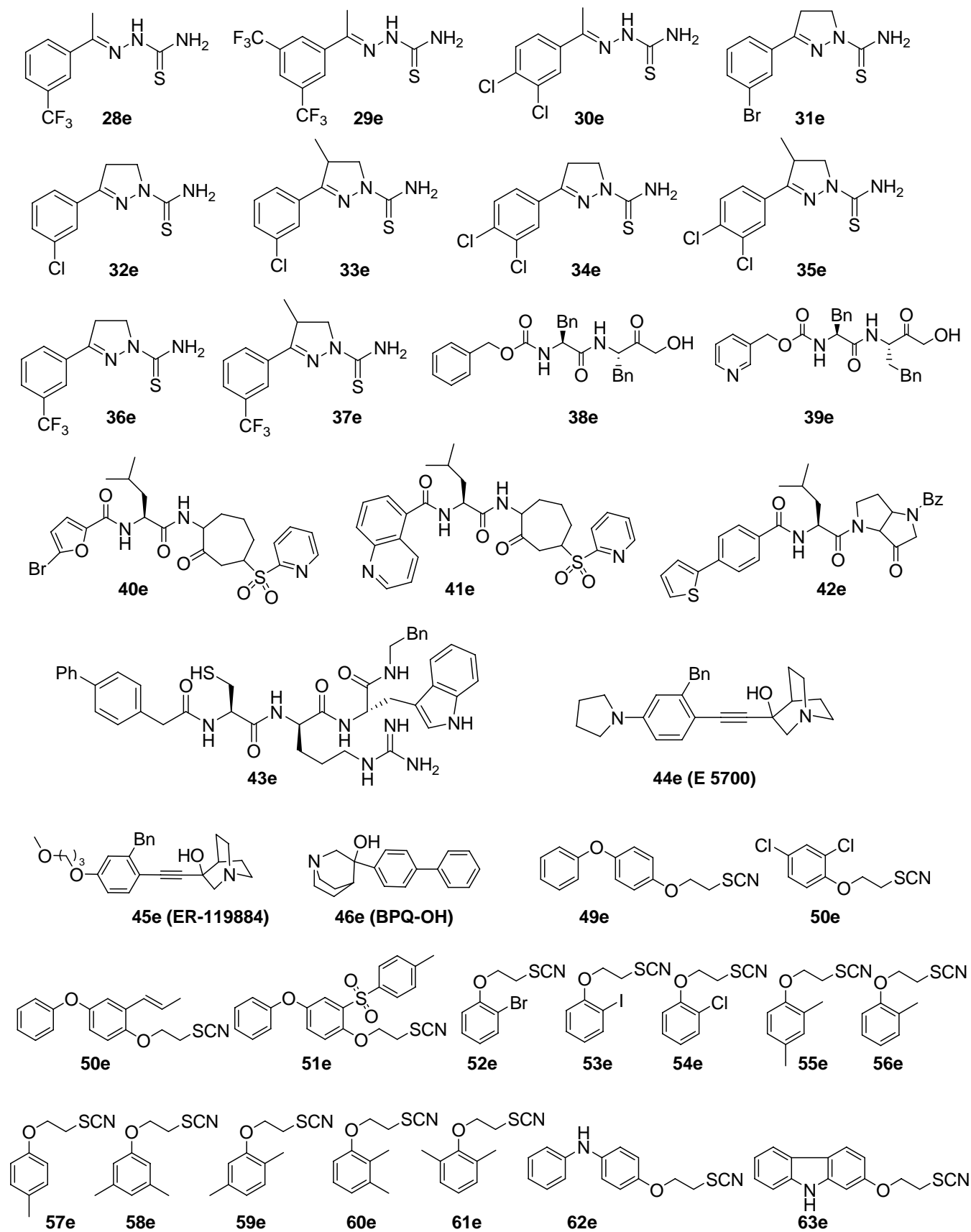
Compounds	Eq.1	Eq. 2
Va2-25	86.72	-35.89
Va2-38	90.43	-52.95
Va5-5b	98.73	60.34
Va6-6b	58.83	-31.97
EICAR	-98.58	-78.77
Mizoribine	-99.42	-79.25
Ribavirin	-98.29	-84.19
Foscarnet	-99.88	-93.84
IMPY	-93.08	-88.51
Citenazone	-94.55	-60.30
Aciclovir	-99.71	-62.74
AIDU	-99.03	99.75
RS-21592	-99.77	-81.55
SKF-23880 A	-81.23	-90.57
Carbodine	-98.29	-79.76
BVDU	-99.39	9.84
-	-	-
Aphidicolin	100.00	78.03
Zalcitabine	-96.10	-82.21
Valaciclovir	-99.79	-83.87
Penciclovir	-99.30	-85.83
Clomethiazole	-98.41	-89.70
Nevanide	-83.99	-57.98
Zonisamide	-93.92	-35.29
LU 2443	-87.25	-32.42
IL-16	-95.20	99.00
Beclamide	-95.08	-24.11
Zebromal	-97.95	95.29
Hierro Girard	-98.43	-90.07
Ferromaltose	-99.57	-55.63
Metipamid	-99.43	26.90
Trometamol	-99.51	-92.45
Theobromine	-99.37	-89.29
Flumethiazide	-99.97	-48.24
WR 2823	-99.56	-78.53
Barbismetylii iodidum	-99.98	-92.61
Basedol	-98.26	-90.55
Mitoguazone	-99.63	-89.59
Metaneophrine	-97.43	-69.38

**Table S4.** *Cont...*

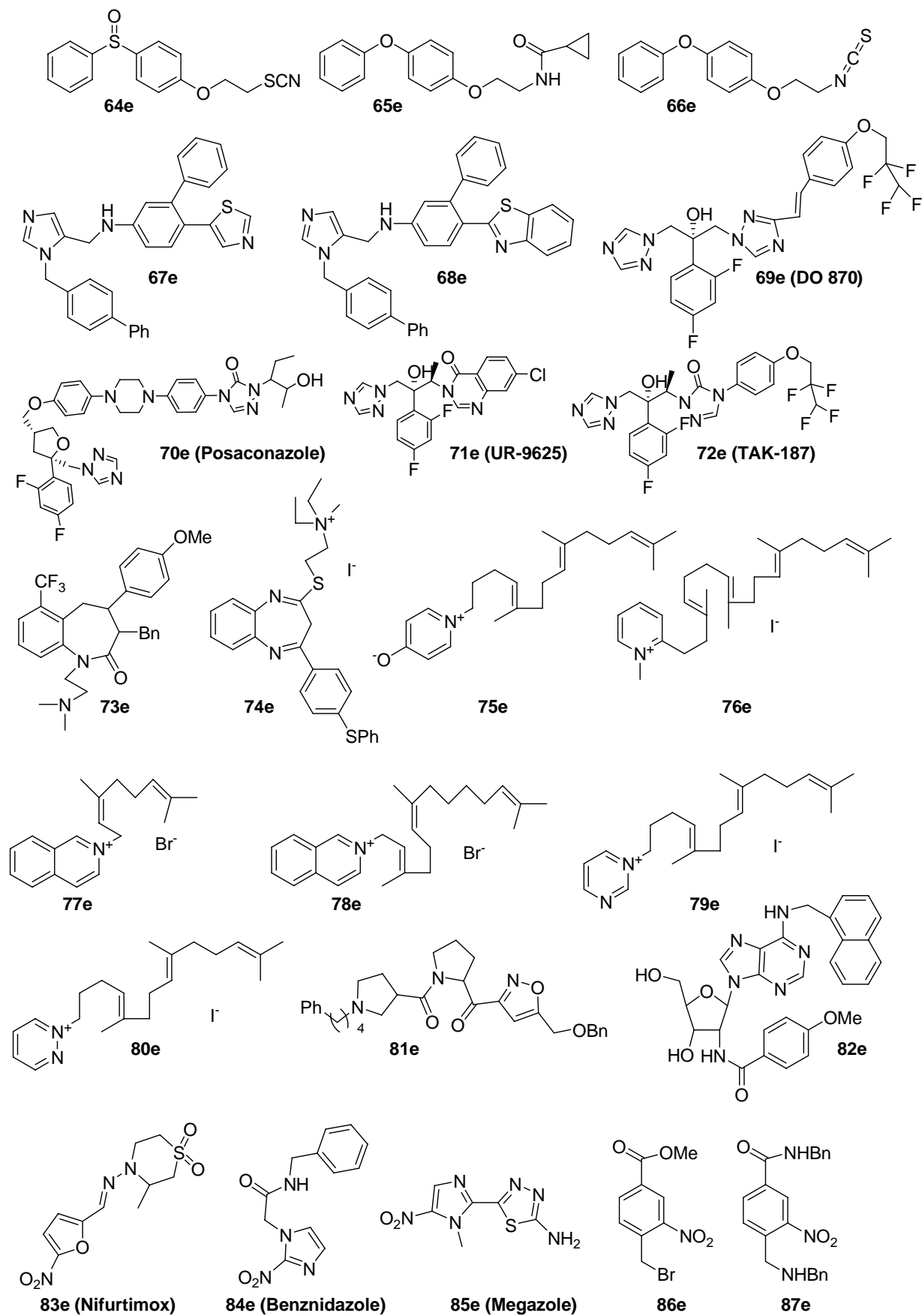
Compounds	Eq.1	Eq. 2
Acebutolol hydrochloride	-92.52	-54.85
Betaxolol	-87.49	-31.71
Bisoprolol fumarate	-96.00	-65.94
Pafenolol	-95.49	-73.12
Talinolol	-95.42	-27.88
Cassaidine	-99.98	43.22
Tioctilate	-40.32	-69.92
Arphamenine A	-93.48	0.56
Homocysteine thiolactone	-94.15	-91.91
Xanthine	-98.74	-88.25
Ethyl glutamate	-98.30	-90.90
Thioctamide	-93.98	-88.74
Beta-sitosterol	-99.75	85.56
Ambucetamide	-87.23	-77.84
Butaverine	-52.59	-32.53
Drofenine	-76.80	-53.13
Fenalamide	-96.14	-69.19
Phenamacide hydrochloride	-65.66	-52.62
Arifen	-99.69	-94.94
Fluopsin C	-98.86	-89.49
Fludalanine	-99.84	-89.07
Cycloserine	-98.19	-91.65
Cryptargol	-99.65	-97.71
Emimycin	-96.22	-89.51
Enheptin-P	-99.35	-88.98
Protoanemonin	-85.16	-79.61
Azaserine	-99.24	-86.80
Contramine 'B	-97.11	-94.57
Iodosil-Tabl	-98.03	99.78
Nitrofurylather	-99.54	-89.87
1441 Cefalexin	-89.63	-4.26
1445 Ampicillin	-90.52	-0.17
1446 Kanamycin A	-99.51	-18.29



**Figure S1.** Molecular structures of anti-trypansomal compounds used in training set



**Figure S1. Cont...**



**Figure S1. Cont...**

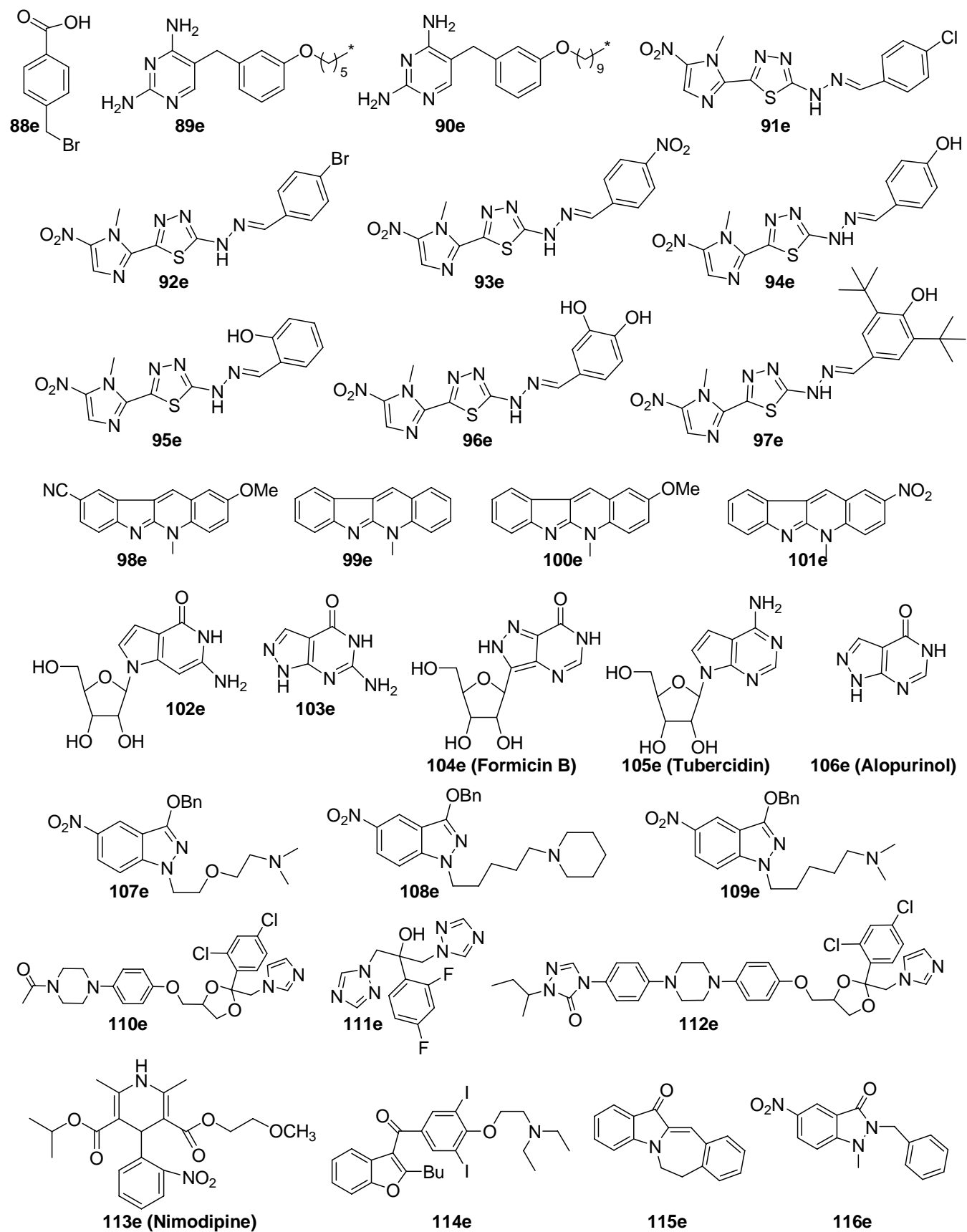
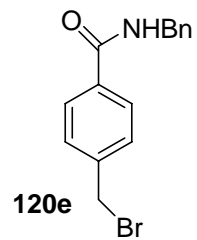
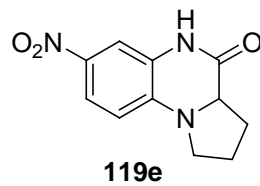
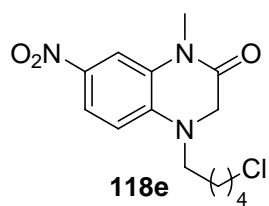
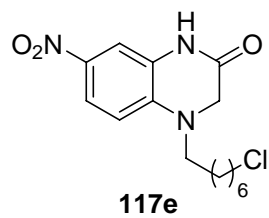


Figure S1. *Cont...*



**Figure S1.** *Cont...*

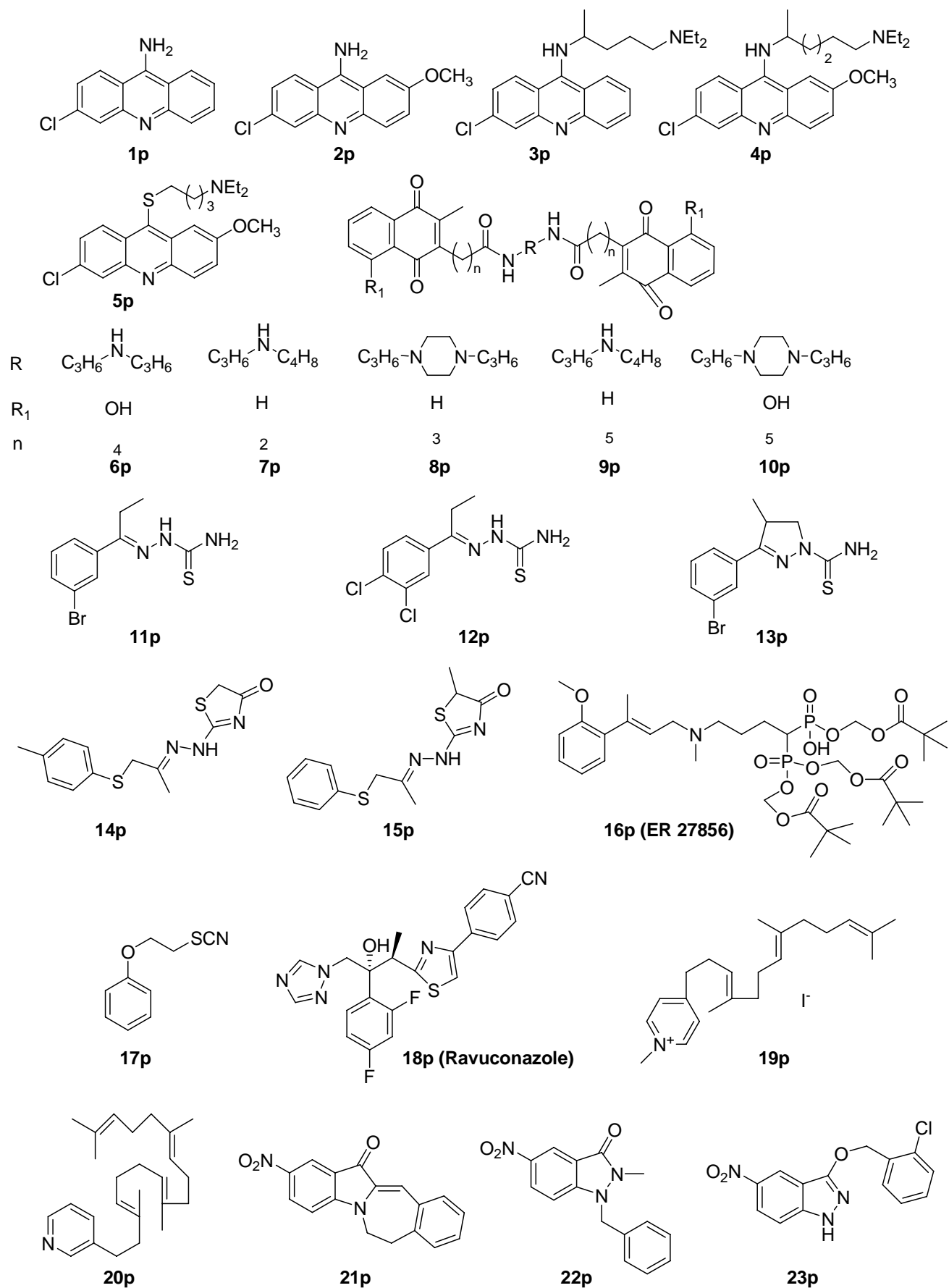


Figure S2. . Molecular structures of anti-trypanosomal compounds used in test set