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,a,b,c* Maria C. Vega,d Miriam Rolon, Yovani Marrero-Ponce, Vladimir V. Kouznetsov, Diego Fernando Amado Torres, Alicia Gómez-Barrio, Alfredo Alvarez Bello, Alina Montero, Francisco Torrens, and Facundo Pérez-Giménez.

aApplied Chemistry Research Center, Faculty of Chemistry-Pharmacy, Central University of Las Villas, Santa Clara, 54830, Villa Clara, Cuba. e-mail: jacgarit@yahoo.es, juancg.22@gmail.com or juancg@uclv.edu.cu
bUnit 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.
cInstitut Universitari de Ciència Molecular, Universitat de València, Edifici d’Instituts de Paterna, P.O. Box 22085, E-46071, València, Spain.
dDepartamento de Parasitología, Facultad de Farmacia, UCM, Pza. Ramón y Cajal s/n, 28040 Madrid.
eUnidad de Investigación de Diseño de Fármacos y Conectividad Molecular, Departamento de Química Física, Facultad de Farmacia, Universitat de València, Spain.
fLaboratorio 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 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 TOpological MOlecular COMputer Design] (Marrero-Ponce and Romero, 2002; Marrero-Ponce, 2003; Marrero Ponce, 2004; Marrero-Ponce et al., 2006c; Casañola-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 paramphistomicides (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 publication (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) $k^{th}$ ($k = 15$) total non-stochastic bond-based quadratic indices, not considering and considering H-atoms in the molecular graph (G) $[q_k(\overline{w})$ and $q_k^H(\overline{w})$, respectively].

2) $k^{th}$ ($k = 15$) total stochastic bond-based quadratic indices, not considering and considering H-atoms in the molecular graph (G) $[s_qk(\overline{w})$ and $s_qk^H(\overline{w})$, respectively].

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

4) $k^{th}$ ($k = 15$) bond-type local (group = heteroatoms: S, N, O) stochastic quadratic indices, not considering and considering H-atoms in the molecular graph (G) $[s_q_{kL}(\overline{w}_E)$ and $s_q_{kL}^H(\overline{w}_E)$, 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’ Tripanosomicidals 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 28º 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 x 10⁶ 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 28º 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:

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

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 56ºC), 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 37ºC in a humidified 5% CO₂ 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 × 10^4 cells/well) in 96-well flat-bottom microplates with 100 µl of RPMI 1640 medium. The cells were allowed to attach for 24 h at 37ºC, 5% CO₂ and the medium was replaced by different concentrations of the drugs in 200 µl of medium, and exposed for another 24 h. Growth controls were also included. Afterwards, a volume 20 µl 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:

Class= -2.77 – 1.73x10^{-1} G_{q0H(\overline{X}E)} + 8.62x10^{-2} G_{q1H(\overline{X})} – 5.94x10^{-4} G_{q4(\overline{X})} \\
+ 7.10x10^{G_{q12L(\overline{X}E)} – 2.32x10^{-2} G_{q1L(\overline{X}E)} – 7.52x10^{-7} G_{q8H(\overline{X})}} \tag{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

Class= -3.54 – 1.04x10^{-2} Ms_{q2(\overline{X})} + 5.27x10^{-3} Ms_{q14L(\overline{X}E)} + 1.54x10^{-2} Ms_{q4(\overline{X})} \\
+ 5.43x10^{-3} Ms_{q2L(\overline{X}E)} + 1.44x10^{-2} Ms_{q0H(\overline{X})} – 1.40x10^{-2} Ms_{q0LH(\overline{X}E)}
\[ -8.01 \times 10^{-3} M_q \Phi_3(\bar{X}) + 4.35 \times 10^{-2} M_q \Phi_{15L}(\bar{X}_E) - 1.43 \times 10^{-1} M_q \Phi_{11L}(\bar{X}_E) \\
+ 9.13 \times 10^{-2} M_q \Phi_L(\bar{X}_E) \] 

(2)

\[ N = 346 \quad \lambda = 0.50 \quad Q_{Total} = 87.57\% \quad MCC = 0.73 \]

\[ D^2 = 4.41 \quad F = 33.69 \quad p < 0.0001 \]

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.**

<table>
<thead>
<tr>
<th>Models</th>
<th>Matthews Corr. Coefficient (C)</th>
<th>Accuracy 'Q_{Total}' (%)</th>
<th>Specificity (%)</th>
<th>Sensitivity 'hit rate' (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Training set</td>
<td>Test set</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eq. 1</td>
<td>0.86</td>
<td>93.35</td>
<td>84.89</td>
<td>98.33</td>
</tr>
<tr>
<td>Eq. 2</td>
<td>0.73</td>
<td>87.57</td>
<td>78.95</td>
<td>87.50</td>
</tr>
<tr>
<td>Eq. 1</td>
<td>90.00</td>
<td>95.74</td>
<td>85.19</td>
<td>100.00</td>
</tr>
<tr>
<td>Eq. 2</td>
<td>69.00</td>
<td>86.17</td>
<td>65.63</td>
<td>91.30</td>
</tr>
</tbody>
</table>

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 ΔP% 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 100µg/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 10µg/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 100µg/mL also showed high values of cytotoxicity (53.20% and 44.06%, respectively). At a concentration of 10µg/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 100µg/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 (ΔP%) according to the obtained models, their antitrypanosomal activity and cytotoxicity at three different concentrations (100, 10, and 1 µg/mL) and antitrypanosomal activity of nifurtimox (reference).

<table>
<thead>
<tr>
<th>Compound</th>
<th>Exp.</th>
<th>ΔP Eq. 1&lt;sup&gt;b&lt;/sup&gt;</th>
<th>ΔP Eq. 2&lt;sup&gt;c&lt;/sup&gt;</th>
<th>%AE (SD)&lt;sup&gt;d&lt;/sup&gt;</th>
<th>%CI&lt;sup&gt;e&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>100µg/mL</td>
<td>10µg/mL</td>
<td>1µg/mL</td>
<td>100µg/mL</td>
</tr>
<tr>
<td>FER10</td>
<td>I</td>
<td>-80.03</td>
<td>-3.17</td>
<td>15.32± 1.4</td>
<td>5.07± 1.28</td>
</tr>
<tr>
<td>FER16</td>
<td>A</td>
<td>-22.93</td>
<td>62.58</td>
<td>86.74± 1.73</td>
<td>78.22± 1.37</td>
</tr>
<tr>
<td>FER19</td>
<td>A</td>
<td>-7.13</td>
<td>69.25</td>
<td>69.19± 1.02</td>
<td>3.17± 0.28</td>
</tr>
<tr>
<td>FER25</td>
<td>I</td>
<td>-79.51</td>
<td>-4.10</td>
<td>59.13± 0.53</td>
<td>42.87± 1.4</td>
</tr>
<tr>
<td>FER26</td>
<td>I</td>
<td>-72.76</td>
<td>1.04</td>
<td>10.39± 2.05</td>
<td>4.10± 2.21</td>
</tr>
<tr>
<td>FER29</td>
<td>I</td>
<td>-99.67</td>
<td>86.76</td>
<td>13.10± 1.22</td>
<td>1.98± 1.48</td>
</tr>
<tr>
<td>FER32</td>
<td>A</td>
<td>21.67</td>
<td>71.30</td>
<td>78.12± 0.71</td>
<td>44.99± 2.52</td>
</tr>
<tr>
<td>FER33</td>
<td>A</td>
<td>19.37</td>
<td>71.68</td>
<td>88.85± 3.44</td>
<td>81.31± 0.76</td>
</tr>
<tr>
<td>FER132</td>
<td>A</td>
<td>-89.49</td>
<td>53.82</td>
<td>72.10±0.28</td>
<td>38.20±2.61</td>
</tr>
<tr>
<td>Nifurtimox</td>
<td>A</td>
<td>99.98</td>
<td>98.39</td>
<td>100±1.49</td>
<td>85.45±2.43</td>
</tr>
</tbody>
</table>

<sup>a</sup>Observed activity.
<sup>b</sup>Results of the classification of compounds obtained from Model 1, DP% = [P(active) - P(inactive)] · 100
<sup>c</sup>Results of the classification of compounds obtained from Model 2, DP% = [P(active) - P(inactive)] · 100
<sup>d</sup>Anti-epimastigotes percentage and standard deviation (SD)
<sup>e</sup>Cytotoxicity percentage
A: active
I: inactive
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


Marrero-Ponce, Y., Romero, V., 2002. TOMOCOMD-CARDD software. TOMOCOMD (TOTopological 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.


Supporting Information

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

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

Table of Contents

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>Classification of active compounds included in the training set using Equations 1 and 2</td>
<td>2</td>
</tr>
<tr>
<td>S2</td>
<td>Classification of inactive compounds included in the training set using Equations 1 and 2</td>
<td>4</td>
</tr>
<tr>
<td>S3</td>
<td>Classification of active compounds included in the test set using Equations 1 and 2</td>
<td>7</td>
</tr>
<tr>
<td>S4</td>
<td>Classification of inactive compounds included in the test set using Equations 1 and 2</td>
<td>8</td>
</tr>
<tr>
<td>S1</td>
<td>Molecular structures of anti-trypanosomal compounds used in training set</td>
<td>9</td>
</tr>
<tr>
<td>S2</td>
<td>Molecular structures of anti-trypanosomal compounds used in test set</td>
<td>14</td>
</tr>
<tr>
<td>Compounds</td>
<td>Eq. 1</td>
<td>Eq. 2</td>
</tr>
<tr>
<td>-----------</td>
<td>-------</td>
<td>-------</td>
</tr>
<tr>
<td>1e</td>
<td>63.0</td>
<td>58.3</td>
</tr>
<tr>
<td>2e</td>
<td>86.5</td>
<td>37.1</td>
</tr>
<tr>
<td>3e</td>
<td>94.7</td>
<td>71.6</td>
</tr>
<tr>
<td>4e</td>
<td>98.1</td>
<td>86.2</td>
</tr>
<tr>
<td>5e</td>
<td>33.6</td>
<td>8.6</td>
</tr>
<tr>
<td>6e</td>
<td>68.5</td>
<td>41.0</td>
</tr>
<tr>
<td>7e</td>
<td>96.5</td>
<td>60.9</td>
</tr>
<tr>
<td>8e</td>
<td>81.0</td>
<td>90.1</td>
</tr>
<tr>
<td>9e</td>
<td>99.9</td>
<td>57.3</td>
</tr>
<tr>
<td>10e</td>
<td>69.0</td>
<td>97.1</td>
</tr>
<tr>
<td>11e</td>
<td>98.3</td>
<td>95.7</td>
</tr>
<tr>
<td>12e</td>
<td>89.6</td>
<td>91.3</td>
</tr>
<tr>
<td>13e</td>
<td>93.8</td>
<td>-93.7</td>
</tr>
<tr>
<td>14e</td>
<td>98.6</td>
<td>38.5</td>
</tr>
<tr>
<td>15e</td>
<td>94.5</td>
<td>-78.0</td>
</tr>
<tr>
<td>16e</td>
<td>99.9</td>
<td>98.7</td>
</tr>
<tr>
<td>17e</td>
<td>99.9</td>
<td>99.2</td>
</tr>
<tr>
<td>18e</td>
<td>100.0</td>
<td>98.6</td>
</tr>
<tr>
<td>19e</td>
<td>99.9</td>
<td>99.6</td>
</tr>
<tr>
<td>20e</td>
<td>99.9</td>
<td>99.5</td>
</tr>
<tr>
<td>21e</td>
<td>99.0</td>
<td>98.7</td>
</tr>
<tr>
<td>22e</td>
<td>94.2</td>
<td>76.3</td>
</tr>
<tr>
<td>23e</td>
<td>92.2</td>
<td>28.7</td>
</tr>
<tr>
<td>24e</td>
<td>94.1</td>
<td>71.4</td>
</tr>
<tr>
<td>25e</td>
<td>93.9</td>
<td>99.9</td>
</tr>
<tr>
<td>25e</td>
<td>100.0</td>
<td>96.0</td>
</tr>
<tr>
<td>27e</td>
<td>99.7</td>
<td>94.0</td>
</tr>
<tr>
<td>28e</td>
<td>99.7</td>
<td>92.4</td>
</tr>
<tr>
<td>29e</td>
<td>100.0</td>
<td>95.4</td>
</tr>
<tr>
<td>30e</td>
<td>74.9</td>
<td>90.3</td>
</tr>
<tr>
<td>31e</td>
<td>63.2</td>
<td>98.8</td>
</tr>
<tr>
<td>32e</td>
<td>64.8</td>
<td>43.1</td>
</tr>
<tr>
<td>33e</td>
<td>59.9</td>
<td>54.1</td>
</tr>
<tr>
<td>34e</td>
<td>82.9</td>
<td>78.3</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Compounds</th>
<th>Eq. 1</th>
<th>Eq. 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>35e</td>
<td>80.3</td>
<td>84.2</td>
</tr>
<tr>
<td>36e</td>
<td>99.8</td>
<td>60.1</td>
</tr>
<tr>
<td>37e</td>
<td>99.7</td>
<td>66.0</td>
</tr>
<tr>
<td>38e</td>
<td>99.9</td>
<td>99.0</td>
</tr>
<tr>
<td>39e</td>
<td>99.9</td>
<td>98.8</td>
</tr>
<tr>
<td>40e</td>
<td>98.3</td>
<td>93.4</td>
</tr>
<tr>
<td>41e</td>
<td>97.8</td>
<td>69.1</td>
</tr>
<tr>
<td>42e</td>
<td>99.4</td>
<td>94.8</td>
</tr>
<tr>
<td>43e</td>
<td>100.0</td>
<td>100.0</td>
</tr>
<tr>
<td>44e</td>
<td>25.5</td>
<td>97.9</td>
</tr>
<tr>
<td>45e</td>
<td>74.7</td>
<td>99.8</td>
</tr>
<tr>
<td>46e</td>
<td>-47.3</td>
<td>91.34</td>
</tr>
<tr>
<td>47e</td>
<td>95.32</td>
<td>96.91</td>
</tr>
<tr>
<td>48e</td>
<td>84.93</td>
<td>22.81</td>
</tr>
<tr>
<td>49e</td>
<td>97.39</td>
<td>99.39</td>
</tr>
<tr>
<td>50e</td>
<td>99.19</td>
<td>98.92</td>
</tr>
<tr>
<td>51e</td>
<td>63.86</td>
<td>-15.01</td>
</tr>
<tr>
<td>52e</td>
<td>61.84</td>
<td>-28.59</td>
</tr>
<tr>
<td>53e</td>
<td>65.14</td>
<td>51.68</td>
</tr>
<tr>
<td>54e</td>
<td>44.20</td>
<td>91.72</td>
</tr>
<tr>
<td>55e</td>
<td>36.70</td>
<td>90.57</td>
</tr>
<tr>
<td>56e</td>
<td>43.34</td>
<td>79.92</td>
</tr>
<tr>
<td>57e</td>
<td>48.37</td>
<td>89.63</td>
</tr>
<tr>
<td>58e</td>
<td>93.21</td>
<td>95.20</td>
</tr>
<tr>
<td>59e</td>
<td>92.04</td>
<td>96.10</td>
</tr>
<tr>
<td>60e</td>
<td>92.11</td>
<td>95.80</td>
</tr>
<tr>
<td>61e</td>
<td>87.82</td>
<td>96.92</td>
</tr>
<tr>
<td>62e</td>
<td>89.76</td>
<td>99.40</td>
</tr>
<tr>
<td>63e</td>
<td>90.75</td>
<td>97.11</td>
</tr>
<tr>
<td>64e</td>
<td>95.10</td>
<td>67.74</td>
</tr>
<tr>
<td>65e</td>
<td>94.70</td>
<td>87.58</td>
</tr>
<tr>
<td>66e</td>
<td>99.92</td>
<td>99.90</td>
</tr>
<tr>
<td>67e</td>
<td>99.98</td>
<td>99.97</td>
</tr>
<tr>
<td>68e</td>
<td>99.99</td>
<td>96.23</td>
</tr>
<tr>
<td>69e</td>
<td>99.96</td>
<td>98.84</td>
</tr>
<tr>
<td>70e</td>
<td>98.51</td>
<td>92.53</td>
</tr>
<tr>
<td>71e</td>
<td>99.95</td>
<td>83.14</td>
</tr>
<tr>
<td>72e</td>
<td>99.96</td>
<td>99.99</td>
</tr>
<tr>
<td>73e</td>
<td>90.32</td>
<td>95.67</td>
</tr>
<tr>
<td>Compounds</td>
<td>Eq.1</td>
<td>Eq.2</td>
</tr>
<tr>
<td>-----------</td>
<td>------</td>
<td>------</td>
</tr>
<tr>
<td>74e</td>
<td>81.94</td>
<td>10.84</td>
</tr>
<tr>
<td>75e</td>
<td>75.92</td>
<td>57.03</td>
</tr>
<tr>
<td>76e</td>
<td>65.11</td>
<td>38.57</td>
</tr>
<tr>
<td>77e</td>
<td>87.17</td>
<td>75.88</td>
</tr>
<tr>
<td>78e</td>
<td>56.46</td>
<td>-23.08</td>
</tr>
<tr>
<td>79e</td>
<td>63.18</td>
<td>-6.57</td>
</tr>
<tr>
<td>80e</td>
<td>99.48</td>
<td>97.76</td>
</tr>
<tr>
<td>81e</td>
<td>99.52</td>
<td>98.36</td>
</tr>
<tr>
<td>82e</td>
<td>74.50</td>
<td>-70.67</td>
</tr>
<tr>
<td>83e</td>
<td>99.98</td>
<td>98.39</td>
</tr>
<tr>
<td>84e</td>
<td>99.41</td>
<td>96.57</td>
</tr>
<tr>
<td>85e</td>
<td>82.11</td>
<td>55.45</td>
</tr>
<tr>
<td>86e</td>
<td>90.07</td>
<td>60.23</td>
</tr>
<tr>
<td>87e</td>
<td>93.93</td>
<td>94.20</td>
</tr>
<tr>
<td>88e</td>
<td>20.63</td>
<td>95.03</td>
</tr>
<tr>
<td>89e</td>
<td>83.05</td>
<td>-6.58</td>
</tr>
<tr>
<td>90e</td>
<td>89.27</td>
<td>38.49</td>
</tr>
<tr>
<td>91e</td>
<td>99.64</td>
<td>40.55</td>
</tr>
<tr>
<td>92e</td>
<td>99.62</td>
<td>97.71</td>
</tr>
<tr>
<td>93e</td>
<td>99.88</td>
<td>11.64</td>
</tr>
<tr>
<td>94e</td>
<td>99.31</td>
<td>13.34</td>
</tr>
<tr>
<td>95e</td>
<td>99.21</td>
<td>8.28</td>
</tr>
<tr>
<td>96e</td>
<td>99.39</td>
<td>7.96</td>
</tr>
<tr>
<td>97e</td>
<td>94.52</td>
<td>18.44</td>
</tr>
<tr>
<td>98e</td>
<td>95.54</td>
<td>82.64</td>
</tr>
<tr>
<td>99e</td>
<td>82.26</td>
<td>74.49</td>
</tr>
<tr>
<td>100e</td>
<td>91.85</td>
<td>73.05</td>
</tr>
<tr>
<td>101e</td>
<td>97.65</td>
<td>82.40</td>
</tr>
<tr>
<td>102e</td>
<td>-26.67</td>
<td>-13.89</td>
</tr>
<tr>
<td>103e</td>
<td>3.98</td>
<td>-82.54</td>
</tr>
<tr>
<td>104e</td>
<td>42.86</td>
<td>-23.50</td>
</tr>
<tr>
<td>105e</td>
<td>69.34</td>
<td>-16.31</td>
</tr>
<tr>
<td>106e</td>
<td>44.20</td>
<td>91.72</td>
</tr>
<tr>
<td>107e</td>
<td>98.84</td>
<td>26.69</td>
</tr>
<tr>
<td>108e</td>
<td>99.72</td>
<td>71.02</td>
</tr>
<tr>
<td>109e</td>
<td>99.36</td>
<td>20.89</td>
</tr>
<tr>
<td>110e</td>
<td>99.43</td>
<td>92.52</td>
</tr>
<tr>
<td>111e</td>
<td>88.20</td>
<td>30.51</td>
</tr>
<tr>
<td>112e</td>
<td>91.31</td>
<td>-18.74</td>
</tr>
</tbody>
</table>
Table S2 Classification of inactive compounds included in the training set using Equations 1 and 2

<table>
<thead>
<tr>
<th>Compounds</th>
<th>Eq.1</th>
<th>Eq. 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Va1-11</td>
<td>82.59</td>
<td>-30.71</td>
</tr>
<tr>
<td>Va1-13</td>
<td>92.32</td>
<td>-52.40</td>
</tr>
<tr>
<td>Va1-15</td>
<td>98.15</td>
<td>55.10</td>
</tr>
<tr>
<td>Va1-31</td>
<td>61.82</td>
<td>-50.77</td>
</tr>
<tr>
<td>Va1-33</td>
<td>83.22</td>
<td>-44.74</td>
</tr>
<tr>
<td>Va2-10</td>
<td>89.80</td>
<td>80.25</td>
</tr>
<tr>
<td>Va2-17</td>
<td>97.84</td>
<td>75.34</td>
</tr>
<tr>
<td>Va2-26</td>
<td>26.65</td>
<td>-49.74</td>
</tr>
<tr>
<td>Va2-27</td>
<td>96.83</td>
<td>64.56</td>
</tr>
<tr>
<td>Va3-3c</td>
<td>-80.66</td>
<td>-44.06</td>
</tr>
<tr>
<td>Va3-3f</td>
<td>-6.87</td>
<td>-22.08</td>
</tr>
<tr>
<td>Va5-9a</td>
<td>96.29</td>
<td>45.00</td>
</tr>
<tr>
<td>Va5-10</td>
<td>96.55</td>
<td>25.86</td>
</tr>
<tr>
<td>Va5-14j</td>
<td>99.31</td>
<td>59.25</td>
</tr>
<tr>
<td>Va5-14g</td>
<td>99.75</td>
<td>92.25</td>
</tr>
<tr>
<td>Va6-5b</td>
<td>93.67</td>
<td>-35.17</td>
</tr>
<tr>
<td>Va6-5d</td>
<td>98.48</td>
<td>74.84</td>
</tr>
<tr>
<td>Va6-9a</td>
<td>83.99</td>
<td>-21.23</td>
</tr>
<tr>
<td>Va6-10a</td>
<td>81.91</td>
<td>6.68</td>
</tr>
<tr>
<td>Vax-6</td>
<td>90.65</td>
<td>-52.76</td>
</tr>
<tr>
<td>Vax-7</td>
<td>92.06</td>
<td>-52.15</td>
</tr>
<tr>
<td>Vax-13</td>
<td>-49.53</td>
<td>-61.64</td>
</tr>
<tr>
<td>Amantadine</td>
<td>-96.14</td>
<td>-51.27</td>
</tr>
<tr>
<td>3-Epistostatin B</td>
<td>-99.42</td>
<td>-82.48</td>
</tr>
<tr>
<td>PALA</td>
<td>-99.82</td>
<td>-81.26</td>
</tr>
<tr>
<td>Thiacetazine</td>
<td>-94.09</td>
<td>-3.14</td>
</tr>
<tr>
<td>TBHQ</td>
<td>-94.09</td>
<td>-40.66</td>
</tr>
<tr>
<td>BL-1743</td>
<td>-95.55</td>
<td>-70.85</td>
</tr>
<tr>
<td>Dimepranol</td>
<td>-98.46</td>
<td>-95.73</td>
</tr>
<tr>
<td>LK-274</td>
<td>-98.68</td>
<td>-95.51</td>
</tr>
<tr>
<td>Ethoxene</td>
<td>-91.88</td>
<td>-88.87</td>
</tr>
<tr>
<td>Kethoxal</td>
<td>-99.21</td>
<td>-96.18</td>
</tr>
<tr>
<td>DANA</td>
<td>-99.51</td>
<td>-52.68</td>
</tr>
<tr>
<td>Moroxidine</td>
<td>-99.24</td>
<td>-94.77</td>
</tr>
<tr>
<td>Brometamine</td>
<td>-98.98</td>
<td>-62.67</td>
</tr>
<tr>
<td>Cloral betaine</td>
<td>-99.89</td>
<td>-81.21</td>
</tr>
<tr>
<td>Triclofos</td>
<td>-99.95</td>
<td>-86.00</td>
</tr>
<tr>
<td>Trichlorourethan</td>
<td>-99.78</td>
<td>-99.54</td>
</tr>
<tr>
<td>Isooral</td>
<td>-99.59</td>
<td>-88.65</td>
</tr>
</tbody>
</table>

Table S2. Cont...

<table>
<thead>
<tr>
<th>Compounds</th>
<th>Eq.1</th>
<th>Eq. 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thiourethane</td>
<td>-97.45</td>
<td>-96.02</td>
</tr>
<tr>
<td>Urethane</td>
<td>-97.86</td>
<td>-96.45</td>
</tr>
<tr>
<td>Etiron</td>
<td>-96.93</td>
<td>-97.15</td>
</tr>
<tr>
<td>Bromobutanol</td>
<td>-99.88</td>
<td>-99.89</td>
</tr>
<tr>
<td>Aleudrin&quot;Beckman&quot;</td>
<td>-99.91</td>
<td>-78.75</td>
</tr>
<tr>
<td>Chlorobutanol</td>
<td>-99.92</td>
<td>-95.17</td>
</tr>
<tr>
<td>Vernelan</td>
<td>-98.73</td>
<td>-89.03</td>
</tr>
<tr>
<td>Agr 449</td>
<td>-97.03</td>
<td>-90.85</td>
</tr>
<tr>
<td>Sedanfactor</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Solucin</td>
<td>-95.79</td>
<td>-91.62</td>
</tr>
<tr>
<td>Cloretate</td>
<td>-99.97</td>
<td>-95.95</td>
</tr>
<tr>
<td>Alcabrol</td>
<td>-98.58</td>
<td>-39.46</td>
</tr>
<tr>
<td>Valerium Paul</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thibault</td>
<td>-93.01</td>
<td>-94.50</td>
</tr>
<tr>
<td>Methonal</td>
<td>-99.66</td>
<td>-95.34</td>
</tr>
<tr>
<td>Nitroinosite</td>
<td>-100.0</td>
<td>67.09</td>
</tr>
<tr>
<td>Cetohexazine</td>
<td>-90.19</td>
<td>-83.43</td>
</tr>
<tr>
<td>Bason</td>
<td>-98.98</td>
<td>-84.76</td>
</tr>
<tr>
<td>Methylpentynol</td>
<td>-94.66</td>
<td>-90.33</td>
</tr>
<tr>
<td>Bromisoval</td>
<td>-99.39</td>
<td>-77.05</td>
</tr>
<tr>
<td>Carbomide</td>
<td>-98.79</td>
<td>-92.82</td>
</tr>
<tr>
<td>Ethchlorvynol</td>
<td>-98.51</td>
<td>-81.43</td>
</tr>
<tr>
<td>Penthichloral</td>
<td>-99.85</td>
<td>-81.14</td>
</tr>
<tr>
<td>Mepentamate</td>
<td>-96.87</td>
<td>-80.58</td>
</tr>
<tr>
<td>Carbavin</td>
<td>-97.46</td>
<td>-83.07</td>
</tr>
<tr>
<td>Phencamid</td>
<td>-89.73</td>
<td>-24.16</td>
</tr>
<tr>
<td>Atrolactamide</td>
<td>-95.92</td>
<td>-46.25</td>
</tr>
<tr>
<td>Tetharbital</td>
<td>-99.15</td>
<td>-74.91</td>
</tr>
<tr>
<td>Brofoxine</td>
<td>-97.80</td>
<td>96.38</td>
</tr>
<tr>
<td>Norantoin</td>
<td>-91.17</td>
<td>-54.19</td>
</tr>
<tr>
<td>Sultaine</td>
<td>-94.99</td>
<td>-53.48</td>
</tr>
<tr>
<td>Albutoin</td>
<td>-93.19</td>
<td>-95.05</td>
</tr>
<tr>
<td>Athotoin</td>
<td>-87.64</td>
<td>-56.44</td>
</tr>
<tr>
<td>Nirvanol</td>
<td>-92.16</td>
<td>-28.50</td>
</tr>
<tr>
<td>Fenaciol</td>
<td>-94.28</td>
<td>-37.79</td>
</tr>
<tr>
<td>Phenetiduride</td>
<td>-88.86</td>
<td>7.09</td>
</tr>
<tr>
<td>Tetrantiol</td>
<td>-86.73</td>
<td>-0.11</td>
</tr>
<tr>
<td>Phenobarbital</td>
<td>-94.68</td>
<td>5.26</td>
</tr>
<tr>
<td>Mesuximide</td>
<td>-90.99</td>
<td>0.35</td>
</tr>
<tr>
<td>Compounds</td>
<td>Eq.1</td>
<td>Eq. 2</td>
</tr>
<tr>
<td>-----------------------</td>
<td>----------</td>
<td>-----------</td>
</tr>
<tr>
<td>Phenylthione</td>
<td>-89.12</td>
<td>-11.37</td>
</tr>
<tr>
<td>Metetoin</td>
<td>-94.44</td>
<td>-32.47</td>
</tr>
<tr>
<td>Mephenytoin</td>
<td>-94.02</td>
<td>-24.22</td>
</tr>
<tr>
<td>Orontsan Fe</td>
<td>-98.74</td>
<td>-84.69</td>
</tr>
<tr>
<td>Cobalt glutamates</td>
<td>-98.43</td>
<td>-86.68</td>
</tr>
<tr>
<td>Ferrocholinate</td>
<td>-99.50</td>
<td>-95.90</td>
</tr>
<tr>
<td>Ferro ascorbas</td>
<td>-99.37</td>
<td>-78.57</td>
</tr>
<tr>
<td>Arecoline</td>
<td>-97.06</td>
<td>-82.09</td>
</tr>
<tr>
<td>Butanolom</td>
<td>-92.08</td>
<td>-94.81</td>
</tr>
<tr>
<td>Cupriaseptol</td>
<td>-95.49</td>
<td>-86.81</td>
</tr>
<tr>
<td>Etamsylate</td>
<td>-98.06</td>
<td>-81.79</td>
</tr>
<tr>
<td>Sango-Stop</td>
<td>-99.57</td>
<td>-82.66</td>
</tr>
<tr>
<td>Carbazochrome</td>
<td>-99.04</td>
<td>74.06</td>
</tr>
<tr>
<td>Sodium sulfonate</td>
<td>-97.99</td>
<td>29.37</td>
</tr>
<tr>
<td>A-Peest</td>
<td>-79.94</td>
<td>-9.82</td>
</tr>
<tr>
<td>Besunide</td>
<td>-99.85</td>
<td>-94.69</td>
</tr>
<tr>
<td>Vitamin B15</td>
<td>-99.66</td>
<td>98.86</td>
</tr>
<tr>
<td>Spiromolactone</td>
<td>-99.80</td>
<td>-73.41</td>
</tr>
<tr>
<td>Chloromethaminofen</td>
<td>-98.20</td>
<td>-71.08</td>
</tr>
<tr>
<td>Amed</td>
<td>-98.03</td>
<td>-94.78</td>
</tr>
<tr>
<td>Guanamine</td>
<td>-98.35</td>
<td>-94.27</td>
</tr>
<tr>
<td>Glycerol</td>
<td>-98.33</td>
<td>-89.54</td>
</tr>
<tr>
<td>Sucinonic acid</td>
<td>-97.73</td>
<td>-88.99</td>
</tr>
<tr>
<td>Pamabron</td>
<td>-98.66</td>
<td>-94.95</td>
</tr>
<tr>
<td>Oxaden</td>
<td>-98.56</td>
<td>-88.83</td>
</tr>
<tr>
<td>Propazolamide</td>
<td>-99.58</td>
<td>-95.78</td>
</tr>
<tr>
<td>Methazolamide</td>
<td>-99.75</td>
<td>-96.25</td>
</tr>
<tr>
<td>Chloromerdrin</td>
<td>-99.91</td>
<td>-75.18</td>
</tr>
<tr>
<td>Theophylline-Merodrin</td>
<td>-99.67</td>
<td>-89.23</td>
</tr>
<tr>
<td>Clofenamide</td>
<td>-99.71</td>
<td>-72.19</td>
</tr>
<tr>
<td>Chloraminophena</td>
<td>-99.83</td>
<td>-76.56</td>
</tr>
<tr>
<td>Butazolamidum</td>
<td>-99.50</td>
<td>-97.88</td>
</tr>
<tr>
<td>Ambuside</td>
<td>-99.61</td>
<td>-26.83</td>
</tr>
<tr>
<td>Merbiurelidin</td>
<td>-99.88</td>
<td>-90.67</td>
</tr>
<tr>
<td>Manna sugar</td>
<td>-99.65</td>
<td>-84.72</td>
</tr>
<tr>
<td>Chlorothiazide</td>
<td>-99.67</td>
<td>-69.32</td>
</tr>
<tr>
<td>Iodothiazide</td>
<td>-99.45</td>
<td>-11.93</td>
</tr>
<tr>
<td>Carzenide</td>
<td>-96.10</td>
<td>-53.25</td>
</tr>
<tr>
<td>Teofillina</td>
<td>-99.37</td>
<td>-89.17</td>
</tr>
<tr>
<td>Siccamid</td>
<td>-99.80</td>
<td>-59.73</td>
</tr>
<tr>
<td>Tectolthiazide</td>
<td>-99.70</td>
<td>-76.67</td>
</tr>
<tr>
<td>Trichlormethiazide</td>
<td>-99.95</td>
<td>80.28</td>
</tr>
<tr>
<td>Propaminnsovet</td>
<td>-94.22</td>
<td>-90.86</td>
</tr>
<tr>
<td>Cystamine</td>
<td>-97.89</td>
<td>-74.17</td>
</tr>
<tr>
<td>Amifostine</td>
<td>-97.95</td>
<td>-95.51</td>
</tr>
<tr>
<td>Adeftonuron</td>
<td>100.00</td>
<td>-72.10</td>
</tr>
<tr>
<td>Batilol</td>
<td>-87.08</td>
<td>-72.56</td>
</tr>
<tr>
<td>Glisalamide</td>
<td>-93.82</td>
<td>-39.45</td>
</tr>
<tr>
<td>Glibutimine</td>
<td>-71.61</td>
<td>-43.86</td>
</tr>
<tr>
<td>RMI 11894</td>
<td>-74.13</td>
<td>-62.41</td>
</tr>
<tr>
<td>Tonyol</td>
<td>-96.63</td>
<td>-86.72</td>
</tr>
<tr>
<td>Tifforminhydrochloride</td>
<td>-98.27</td>
<td>-91.14</td>
</tr>
<tr>
<td>Etoformin hydrochloride</td>
<td>-98.47</td>
<td>-99.09</td>
</tr>
<tr>
<td>Benfosformin</td>
<td>-99.60</td>
<td>-44.01</td>
</tr>
<tr>
<td>Phenformin hydrochloride</td>
<td>-92.81</td>
<td>-52.40</td>
</tr>
<tr>
<td>Mebenformin</td>
<td>-95.85</td>
<td>-56.03</td>
</tr>
<tr>
<td>Glycypyramide</td>
<td>-98.82</td>
<td>-65.52</td>
</tr>
<tr>
<td>Glyprothiazol</td>
<td>-94.10</td>
<td>-85.19</td>
</tr>
<tr>
<td>Phenbutamide</td>
<td>-93.64</td>
<td>-87.57</td>
</tr>
<tr>
<td>Carbutamide</td>
<td>-95.36</td>
<td>-93.53</td>
</tr>
<tr>
<td>Metasulfafulbutylarbamide</td>
<td>-95.49</td>
<td>-92.67</td>
</tr>
<tr>
<td>Centipiperalone</td>
<td>-88.26</td>
<td>-38.81</td>
</tr>
<tr>
<td>Glybuzole</td>
<td>-96.72</td>
<td>-52.74</td>
</tr>
<tr>
<td>SPC-703</td>
<td>-91.90</td>
<td>-84.93</td>
</tr>
<tr>
<td>Tolpyramide</td>
<td>-89.72</td>
<td>-87.67</td>
</tr>
<tr>
<td>Glybuthiazol</td>
<td>-97.62</td>
<td>-72.97</td>
</tr>
<tr>
<td>Thiazanol</td>
<td>-97.21</td>
<td>-94.00</td>
</tr>
<tr>
<td>Tolbutamide</td>
<td>-34.79</td>
<td>-91.67</td>
</tr>
<tr>
<td>Hypoglycin B</td>
<td>-98.29</td>
<td>-32.85</td>
</tr>
<tr>
<td>Anticoman</td>
<td>-97.97</td>
<td>-93.08</td>
</tr>
<tr>
<td>Furfurdurea</td>
<td>-39.71</td>
<td>-33.51</td>
</tr>
<tr>
<td>Compounds</td>
<td>Eq.1</td>
<td>Eq.2</td>
</tr>
<tr>
<td>-----------------------</td>
<td>-------</td>
<td>-------</td>
</tr>
<tr>
<td>Glymidine sodium</td>
<td>-98.18</td>
<td>-3.55</td>
</tr>
<tr>
<td>Cloreexamide</td>
<td>-97.31</td>
<td>-63.72</td>
</tr>
<tr>
<td>Glysobuzole</td>
<td>-96.26</td>
<td>-92.79</td>
</tr>
<tr>
<td>Glypinamide</td>
<td>-98.54</td>
<td>-78.21</td>
</tr>
<tr>
<td>Ag 307</td>
<td>-98.51</td>
<td>-78.83</td>
</tr>
<tr>
<td>Bromobromide</td>
<td>-96.17</td>
<td>-92.97</td>
</tr>
<tr>
<td>Mebetide</td>
<td>-90.12</td>
<td>-70.26</td>
</tr>
<tr>
<td>Minoxidil</td>
<td>-97.27</td>
<td>-91.71</td>
</tr>
<tr>
<td>Hemedin</td>
<td>-96.17</td>
<td>-92.97</td>
</tr>
<tr>
<td>Penbutatin</td>
<td>-99.85</td>
<td>-95.12</td>
</tr>
<tr>
<td>Guanoctine</td>
<td>-99.43</td>
<td>-91.11</td>
</tr>
<tr>
<td>hydrochloride</td>
<td>-94.16</td>
<td>96.30</td>
</tr>
<tr>
<td>Guanisouquene sulfate</td>
<td>-98.51</td>
<td>-37.72</td>
</tr>
<tr>
<td>Flutonidine</td>
<td>-92.73</td>
<td>-65.23</td>
</tr>
<tr>
<td>hydrochloride</td>
<td>-82.64</td>
<td>-49.33</td>
</tr>
<tr>
<td>Nicopholine</td>
<td>-77.36</td>
<td>-52.22</td>
</tr>
<tr>
<td>Ganglefene</td>
<td>-99.74</td>
<td>-97.02</td>
</tr>
<tr>
<td>hydrochloride</td>
<td>-99.66</td>
<td>-91.80</td>
</tr>
<tr>
<td>Quateron</td>
<td>-97.23</td>
<td>-92.12</td>
</tr>
<tr>
<td>Pancuronium bromide</td>
<td>-99.68</td>
<td>-96.95</td>
</tr>
<tr>
<td>Plegarol</td>
<td>-99.43</td>
<td>-96.62</td>
</tr>
<tr>
<td>Gaplagent</td>
<td>-99.56</td>
<td>-96.23</td>
</tr>
<tr>
<td>Oxaditox</td>
<td>-99.74</td>
<td>-97.02</td>
</tr>
<tr>
<td>Agentit</td>
<td>-99.66</td>
<td>-91.80</td>
</tr>
<tr>
<td>Dimecolonium iodide</td>
<td>-97.23</td>
<td>-92.12</td>
</tr>
<tr>
<td>Methylene iodide</td>
<td>-99.66</td>
<td>-91.80</td>
</tr>
<tr>
<td>Hexamethonium iodide</td>
<td>-98.49</td>
<td>-83.85</td>
</tr>
<tr>
<td>Trepirium iodide</td>
<td>-98.86</td>
<td>-68.44</td>
</tr>
<tr>
<td>Tiamethonium iodide</td>
<td>-99.67</td>
<td>-87.39</td>
</tr>
<tr>
<td>Iodide</td>
<td>-99.73</td>
<td>-97.72</td>
</tr>
<tr>
<td>Imekhin</td>
<td>-98.71</td>
<td>-58.94</td>
</tr>
<tr>
<td>Dimecamine</td>
<td>-99.74</td>
<td>-91.63</td>
</tr>
<tr>
<td>Penhexamine</td>
<td>-90.04</td>
<td>-37.69</td>
</tr>
<tr>
<td>Leptodactyline</td>
<td>-98.75</td>
<td>-93.08</td>
</tr>
<tr>
<td>Tetrammonium iodide</td>
<td>-98.86</td>
<td>-68.44</td>
</tr>
<tr>
<td>Iodide</td>
<td>-98.67</td>
<td>-87.39</td>
</tr>
<tr>
<td>Leptodactyline</td>
<td>-96.80</td>
<td>-33.07</td>
</tr>
</tbody>
</table>
### Table S2.  
Continued...

<table>
<thead>
<tr>
<th>Compounds</th>
<th>Eq. 1</th>
<th>Eq. 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Etafedrine</td>
<td>-88.38</td>
<td>-50.23</td>
</tr>
<tr>
<td>hydrochloride</td>
<td>-94.55</td>
<td>-44.88</td>
</tr>
<tr>
<td>Isorenaline</td>
<td>-88.63</td>
<td>-47.75</td>
</tr>
<tr>
<td>Isofenefrine</td>
<td>-79.58</td>
<td>-48.10</td>
</tr>
<tr>
<td>Phenylethanolamine</td>
<td>-96.35</td>
<td>-46.32</td>
</tr>
<tr>
<td>Adrenaline borate</td>
<td>-61.90</td>
<td>-56.61</td>
</tr>
<tr>
<td>Amfetamine</td>
<td>-82.82</td>
<td>-55.21</td>
</tr>
<tr>
<td>Norephedrine</td>
<td>-90.25</td>
<td>-57.43</td>
</tr>
<tr>
<td>Metaraminol</td>
<td>-91.90</td>
<td>-63.53</td>
</tr>
<tr>
<td>Syneprine</td>
<td>-95.35</td>
<td>-54.42</td>
</tr>
<tr>
<td>Corbadrine</td>
<td>-99.86</td>
<td>-79.67</td>
</tr>
<tr>
<td>Nibroxane</td>
<td>-99.84</td>
<td>-97.41</td>
</tr>
<tr>
<td>Alafosfalin</td>
<td>-99.65</td>
<td>-64.96</td>
</tr>
<tr>
<td>Furaguanidine</td>
<td>-94.07</td>
<td>-85.82</td>
</tr>
<tr>
<td>Sulfanilamide</td>
<td>-97.11</td>
<td>29.68</td>
</tr>
<tr>
<td>Parabotrine</td>
<td>-92.65</td>
<td>-84.77</td>
</tr>
<tr>
<td>Allicin</td>
<td>-99.88</td>
<td>22.37</td>
</tr>
<tr>
<td>1442 Streptomycine</td>
<td>83.05</td>
<td>73.11</td>
</tr>
<tr>
<td>Azirinomycin</td>
<td>-95.54</td>
<td>-88.18</td>
</tr>
<tr>
<td>1444 Gentamicin A</td>
<td>-99.32</td>
<td>-34.84</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Compounds</th>
<th>Eq. 1</th>
<th>Eq. 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1p</td>
<td>84.32</td>
<td>78.61</td>
</tr>
<tr>
<td>2p</td>
<td>92.80</td>
<td>82.28</td>
</tr>
<tr>
<td>3p</td>
<td>88.31</td>
<td>65.91</td>
</tr>
<tr>
<td>4p</td>
<td>99.17</td>
<td>88.68</td>
</tr>
<tr>
<td>5p</td>
<td>97.23</td>
<td>41.36</td>
</tr>
<tr>
<td>6p</td>
<td>99.93</td>
<td>98.03</td>
</tr>
<tr>
<td>7p</td>
<td>99.94</td>
<td>97.88</td>
</tr>
<tr>
<td>8p</td>
<td>99.92</td>
<td>99.33</td>
</tr>
<tr>
<td>9p</td>
<td>99.94</td>
<td>98.08</td>
</tr>
<tr>
<td>10p</td>
<td>99.90</td>
<td>99.73</td>
</tr>
<tr>
<td>11p</td>
<td>55.44</td>
<td>99.15</td>
</tr>
<tr>
<td>12p</td>
<td>78.78</td>
<td>90.86</td>
</tr>
<tr>
<td>13p</td>
<td>58.27</td>
<td>99.43</td>
</tr>
<tr>
<td>14p</td>
<td>87.94</td>
<td>-31.60</td>
</tr>
<tr>
<td>15p</td>
<td>86.34</td>
<td>-45.76</td>
</tr>
<tr>
<td>16p</td>
<td>83.05</td>
<td>73.11</td>
</tr>
<tr>
<td>17p</td>
<td>34.32</td>
<td>74.36</td>
</tr>
<tr>
<td>18p</td>
<td>98.74</td>
<td>95.50</td>
</tr>
<tr>
<td>19p</td>
<td>55.58</td>
<td>29.18</td>
</tr>
<tr>
<td>20p</td>
<td>85.33</td>
<td>61.55</td>
</tr>
<tr>
<td>21p</td>
<td>98.05</td>
<td>89.12</td>
</tr>
<tr>
<td>22p</td>
<td>95.66</td>
<td>37.25</td>
</tr>
<tr>
<td>23p</td>
<td>99.23</td>
<td>64.85</td>
</tr>
</tbody>
</table>
### Table S4. Classification of inactive compounds included in the test set using Equations 1 and 2

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

### Table S4. Cont...

<table>
<thead>
<tr>
<th>Compounds</th>
<th>Eq.1</th>
<th>Eq. 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acebutolol hydrochloride</td>
<td>-92.52</td>
<td>-54.85</td>
</tr>
<tr>
<td>Betaxolol</td>
<td>-87.49</td>
<td>-31.71</td>
</tr>
<tr>
<td>Bisoprolol</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fumarate</td>
<td>-96.00</td>
<td>-65.94</td>
</tr>
<tr>
<td>Pafenolol</td>
<td>-95.49</td>
<td>-73.12</td>
</tr>
<tr>
<td>Talinolol</td>
<td>-95.42</td>
<td>-27.88</td>
</tr>
<tr>
<td>Cassaudine</td>
<td>-99.98</td>
<td>43.22</td>
</tr>
<tr>
<td>Tiocitrate</td>
<td>-40.32</td>
<td>-69.92</td>
</tr>
<tr>
<td>Arphenamine A</td>
<td>-93.48</td>
<td>0.56</td>
</tr>
<tr>
<td>Homocysteine</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thiocitrate</td>
<td>-94.15</td>
<td>-91.91</td>
</tr>
<tr>
<td>Xanthine</td>
<td>-98.74</td>
<td>-88.25</td>
</tr>
<tr>
<td>Ethyl glutamate</td>
<td>-98.30</td>
<td>-90.90</td>
</tr>
<tr>
<td>Thiocitamid</td>
<td>-93.98</td>
<td>-88.74</td>
</tr>
<tr>
<td>Beta-sitosterol</td>
<td>-99.75</td>
<td>85.56</td>
</tr>
<tr>
<td>Ambucetamide</td>
<td>-87.23</td>
<td>-77.84</td>
</tr>
<tr>
<td>Butaverine</td>
<td>-52.59</td>
<td>-32.53</td>
</tr>
<tr>
<td>Drofenine</td>
<td>-76.80</td>
<td>-53.13</td>
</tr>
<tr>
<td>Fenalamide</td>
<td>-96.14</td>
<td>-69.19</td>
</tr>
<tr>
<td>Phenazepamide</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hydrochloride</td>
<td>-65.66</td>
<td>-52.62</td>
</tr>
<tr>
<td>Arifen</td>
<td>-99.69</td>
<td>-94.94</td>
</tr>
<tr>
<td>Fluopsin C</td>
<td>-98.86</td>
<td>-89.49</td>
</tr>
<tr>
<td>Fludalanine</td>
<td>-99.84</td>
<td>-89.07</td>
</tr>
<tr>
<td>Cycloserine</td>
<td>-98.19</td>
<td>-91.65</td>
</tr>
<tr>
<td>Cryptargol</td>
<td>-99.65</td>
<td>-97.71</td>
</tr>
<tr>
<td>Emimycin</td>
<td>-96.22</td>
<td>-89.51</td>
</tr>
<tr>
<td>Enheptin-P</td>
<td>-99.35</td>
<td>-88.98</td>
</tr>
<tr>
<td>Protoanemonin</td>
<td>-85.16</td>
<td>-79.61</td>
</tr>
<tr>
<td>Azaserine</td>
<td>-99.24</td>
<td>-86.80</td>
</tr>
<tr>
<td>Contramine 'B</td>
<td>-97.11</td>
<td>-94.57</td>
</tr>
<tr>
<td>Iodosil Tabl</td>
<td>-98.03</td>
<td>99.78</td>
</tr>
<tr>
<td>Nitrofurylether</td>
<td>-99.54</td>
<td>-89.87</td>
</tr>
<tr>
<td>1441 Cefalexin</td>
<td>-89.63</td>
<td>-4.26</td>
</tr>
<tr>
<td>1445 Ampicillin</td>
<td>-90.52</td>
<td>-0.17</td>
</tr>
<tr>
<td>1446 Kanamycin A</td>
<td>-99.51</td>
<td>-18.29</td>
</tr>
</tbody>
</table>
Figure S1. Molecular structures of anti-trypanosomal 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