12th International Electronic Conference on Synthetic Organic Chemistry (ECSOC-12) 1-30 November 2008
http://www.usc.es/congresos/ecsoc/12/ECSOC12.htm & http://www.mdpi.org/ecsoc-12

[C0010]

Antiprotozoan Lead Discovery by Aligning *Dry* and *Wet* Screening: Prediction, Synthesis, and Biological Assay of Novel Quinoxalinones

Miriam A. Martins Alho,[τ] Yovani Marrero-Ponce[†‡*] Alfredo Meneses-Marcel,[†] Yanetsy Machado Tugores,[†] Alina Montero-Torres,[†] Facundo Pérez-Giménez,[‡] Alicia Gómez-Barrio,[ξ] Juan J. Nogal,[ξ] Rory N. García-Sánchez,[ξ,ζ] María Celeste Vega,[ξ] Miriam Rolón,[ξ] Antonio R. Martínez-Fernández,[ξ] José A. Escario,[ξ] Norma Rivera,[ŧ] Froylán Ibarra-Velarde,[τ] Mónica Mondragón,[ŧ] Ricardo Mondragón,[ŧ] Roberto Chicharro,[§] Vicente J. Arán[ℓ***]**

[τ]CIHIDECAR (CONICET), Departamento de Química Orgánica, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, C1428EGA Buenos Aires, Argentina.

[†]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.

[‡]Unidad de Investigación de Diseño de Fármacos y Conectividad Molecular, Departamento de Química Física, Facultad de Farmacia, Universitat de València, Spain.

[ξ]Departamento de Parasitología, Facultad de Farmacia, Universidad Complutense, 28040 Madrid, Spain.

[ζ]Laboratorio de Investigación de Productos Naturales Antiparasitarios de la Amazonía, Universidad Nacional de la Amazonía Peruana, Pasaje Los Paujiles s/n, A A.H H Nuevo San Lorenzo, San Juan Bautista, Iquitos, Perú.

[ŧ]Departamento de Bioquímica, Centro de Investigaciones y Estudios Avanzados del IPN. Av. Instituto Politécnico Nacional No 2508. Col. San Pedro Zacatenco. México DF 07360.

[τ]Department of Parasitology, Faculty of Veterinarian Medicinal and Zootecnic, UNAM, Mexico, D.F. 04510, Mexico.

[§]Instituto de Química Orgánica General, CSIC, c/ Juan de la Cierva 3, 28006 Madrid, Spain.

[ℓ]Instituto de Química Médica, CSIC, c/ Juan de la Cierva 3, 28006 Madrid, Spain.

[*]***To whom correspondence should be addressed*:**


Fax: 53-42-281130 [or 53-42-281455] (Cuba) and 963543156 (València)


Phone: 53-42-281192 [or 53-42-281473] (Cuba) and 963543156 (València)


Cell: 610028990


***e*-mail**: ymarrero77@yahoo.es; ymponce@gmail.com or yovanimp@qf.uclv.edu.cu


URL: http://www.uv.es/yoma/

****(contact for biological assays: escario@farm.ucm.es**
*****(contact for chemical methods: vjaran@iqm.csic.es**

ABSTRACT

Protozoan parasites have been one of the most significant public health problems for centuries and several of human infections causes by them are globally massive in their impact. The most of the current drugs used to treat these illness are decades old and have many limitations, including the emergence of drug resistance, severe side-effects, low-to-medium efficacy, parenteral mode of administration, price, etc. These drugs have been largely neglected for drug development because they affect poor people in poor regions of the world where there is a small market for this kind of drugs. Therefore, nowdays there is a pressing need for identifying and developing new drug-based antiprotozoan therapies. In an effort to overcome this problem, the main purpose of this study is to develop a QSARs-based ensemble classifier for antiprotozoan drug-like compounds from a heterogeneous series of compounds. Here, we use some of the ***TOMOCOMD-CARDD*** molecular descriptors and linear discriminat analisis (LDA) to derive *individual* linear classification functions in order to discriminate between antiprotozoan and nonantiprotozoan compounds, and so as to enable computational screening from virtual combinatorial datasets and/or existing drugs already approved. All studies were carried out taken into account the OECD principle in order for characterizing every obtained QSARs. In first time, a wide-spectrum *benchmark* database of 680 organic chemicals having great structural variability, 254 of them antiprotozoan agents and 426 compounds having other clinical uses, was analyzed and presented as a helpful tool, not only for theoretical chemists but also for other researchers in this area. This series of compounds was processed by a *k*-means cluster analysis in order to design training and predicting sets. In total, seven discriminant functions were obtained, by using the whole set of atom-based linear indices. All the LDA-based QSAR models show accuracies above 85% in the training set and values of Matthews correlation coefficients (*C*) varying from 0.70-0.86. The external validation set shows globally rather-good classifications around 80% (92.05% for best equation). Later, we developed a multi-agent QSAR classification system, in which the individual QSAR outputs are the inputs of the aforementioned fusion approach. Finally, the fusion model was used for the identification of a novel generation of lead-like antiprotozoans by using ligand-based virtual screening of small-molecules 'available' (with synthetic feasibility) in our 'in-house' library. A new molecular *subsystem* (quinoxalinones) was then theoretically selected like promising lead series, which were subsequently synthesized, structurally characterized, and experimentally assayed using an *in vitro* screening that take into consideration a battery of four parasite-based assays. The chemicals **11**(**12**) and **16** are the most active (*hits*) against apicomplexa (sporozoa) and mastigophora (flagellata) *subphylum* parasites, respectively. Both compounds had shown rather good activities in the every protozoan *in vitro* panel and they didn't depict unspecific cytotoxicity to macrophages. This result opens a door to a virtual study considering a higher variability of the structural core already evaluated, as well as of other chemicals not included in this study. We conclude that the approach described here seems to be a promising esamble QSAR-clasifier for the molecular discovery of novel classes of broad –antiprotozoan– spectrum drugs, which may meet the dual challenges posed by drug-resistant parasites and the rapid progression of protozoan illnesses.

Keywords: *In silico* Study, *TOMOCOMD-CARDD* Software, Non-Stochastic and Stochastic Linear Indices, Classification Model, Learning Machine-based QSAR, Antiprotozoan Database, *In vitro* Assay, Antimalarial, Antitrypanosomal, Antotoxoplasma, Antitrichomonas, Cytotocicity.

Running head: *Antiprotozoal Lead Discovery by Aligning Dry and Wet Screening …*

Introduction

Diseases caused by tropical parasites affect hundreds of millions of people worldwide and it concern many tropical and subtropical regions of the world.[1] In fact, parasitic diseases have been one of the most significant public health problems for centuries and now result in noteworthy mortality and devastating social and economic consequences. The parasites include in *phylum protozoa* are the most important pathogens and several of human infections cause by them are globally massive in their impact. For instance, malaria (*Plasmodium* spp.),[2] leishmaniasis (*Leishmania* spp.),[3] trypanosomiasis (*T. brucei* [sleeping sickness][4] and *T. cruzi* [Chagas disease][5]) as well as giardiasis[6]/amebiasis[7] (*Giardia lamblia*/*Entamoeba histolytica*) are among the main neglected parasitic diseases with great social impact.[8] Trichomoniasis, one of the most common sexually transmitted diseases (with around 120 million worldwide suffering from vaginitis every year) caused by the flagellate protozoa *Trichomonas vaginalis*, is increasingly recognized as an important infection in women and men.[9] Other serious disease caused by a related apicomplexan parasite, *Toxoplasma gondii*, takes more and more relevance in immunocompromised patients, such as patients with transplants, cancer, or AIDS, and in congenitally infected infants.[10]

Although protozoa agents are rather common and familiar to most scientists, the most of the current drugs used to treat these illness are decades old and have many limitations, including the emergence of drug resistance, severe side-reactions (toxicity), low-to-medium efficacy, parenteral mode of administration, price and others important inconveniences.[11] These drawbacks of the current antiprotozoan chemotherapy make the search for new drugs urgently needed. However, these drugs have been largely neglected for drug development because they affect poor people in poor regions of the

world where there is a small market for this kind of drugs, particularly in today's post-merger climate.

Nevertheless, the search for antiprotozoan compounds is now on the desktop of medicinal chemists and great efforts to reinvigorate the drug development pipeline for these diseases are being addressed by new consortia of scientists from academia and industry, which is driven in large part by support from major philanthropies.[1] More recently and by using a whole-organism screening of compound libraries containing drugs already approved for human use (with other therapeutic use, but 'off-label' like antiparasitic efficacy), a few *hits* were identified in diversity screens against *T. brucei*, *P. falciparum* and *leishmania*.[12-15] In this "*trial-and-error"* search for antiprotozoan drug-like compounds a lot of chemicals had to be experimentally screened (>15,000) and the efficacy of this process was very low, yielding only 3 (and 20 additional in a second study), 19, and 40 know drugs with efficacy equal to or greater than that of the currently drugs used as leishmania-, malaria- or trypanosoma-reference (control) compound, respectively.[12-15] In addition to the low efficiency of this type of drug discovery landscape, the usually *expensive* and *time consuming* of this kind of search protocol, to impose on us the necessity for development of an alternative (more *rational*) techniques to classical -*trial and error*- screenings, highlights the need for a "sea change" in the drug discovery paradigm.[16] In order to reduce costs, pharmaceutical companies have to find new technologies to the search of new chemical entities (NCE),[17] where an *in silico* 'virtual' world of data, analysis, hypothesis and design that reside inside a computer as well as ligand-based computational screening can be seen like an adequate alternative to the 'real' world of synthesis and screening of compounds in the laboratory. By this means, ''*the expensive commitment to actual synthesis and bioassay is made only after exploring the initial concepts with computational models*

and screens."[18, 19] *In silico* screening is now incorporated in all areas of lead discovery; from target identification and library design, to *hit* analysis and compound profiling.[20] These types of diversity *in silico* screens open up many new avenues for lead discovery and optimization, including the potential to explore natural-product libraries that have so far been largely untapped.[21] This theoretical(*dry*)-to-experimental(*wet*) integration procedure will be used here in order to find predictive models that permit the 'rational' identification of new antiprotozoan drug-like compounds.

Background-Review of *TOMOCOMD-CARDD* Method in Drug Discovery for Parasitic Diseases: *Meeting the Challenge*. In addition to above comment, also there is a widely perceived need for alternative non-animal methods for the biological-assays, ADME and hazard (risk) assessment of chemicals. (Quantitative) structure activity relationships [(Q)SARs] are now being increasingly viewed as one of the most cost effective alternatives to estimate ecological and health effects of chemicals. (Q)SAR[22] predictions have the potential to save time and money as well as minimize the use of animal testing.[23]

Therefore, some of our research teams, previously, have reported several antimicrobial-cheminformatic studies to driven the selection of novel chemicals as promising NCEs. In these studies, the ***TOMOCOMD-CARDD*** **(**acronym of ***TO***pological ***MO***lecular ***COM***puter ***D***esign ***C****omputer-****A****ided –****R****ational– ****D****rug ****D****esign*) method[24] and linear discriminant analysis (LDA),[25] mainly, have been used in order to parameterize every molecule in database and for developing classification functions, respectively. LDA is one of most important and simple (supervise, linear and parametric) patter recognition techniques that can be use to determine which variables discriminate between two or more naturally occurring groups (it is used as either a hypothesis testing or exploratory method-data mining).[25, 26] At present, LDA has

become an significant statistical tool and is rather use in chemometric analysis and drug design studies.[19, 27-29] ***TOMOCOMD-CARDD*** approach is a novel scheme to the rational –*in silico*- molecular design and to QSAR/QSPR.[30-36] It calculates several new families of 2D, 3D-Chiral (2.5) and 3D (geometric and topographic) non-stochastic and (simple and doble) stochastic (as well as canonical their forms) atom- and bond-based molecular descriptors (MDs) based on algebraic theory and discrete mathematic. They are denominate quadratic, linear and bilinear indices and have been defined in analogy to the quadratic, linear and bilinear mathematical maps.[30-36] These approaches describe changes in the electron distribution with time throughout the molecular backbone and they have been successfully employed in the prediction of several physical, physicochemical, chemical biological and pharmacokinetical properties of organic compounds.[37-53] Besides, these indices have been extended to considering three-dimensional features of small/medium-sized molecules based on the *trigonometric 3D-chirality correction factor approach.*[54, 55] In fact, in recent works, we had obtained very promising results when stochastic and non-stochastic 3D-chiral (2.5) quadratic, linear and bilinear indices were applied to three of the most commonly used chiral data sets.[56-59] Recently, our research group reported several classification-based QSAR models, which have been permit the *in silico* discovery of new lead antimicrobial compounds. For instance, the ***TOMOCOMD-CARDD*** strategy has been used for the selection of novel molecular *subsystems* having a desired activity against *Trichomonas vaginalis*.[36, 60, 61] It was also successfully applied to the virtual (computational) screening of novel anthelmintic compounds, which were then synthesized and evaluated *in vivo* on *Fasciola hepatica*.[62, 63] Studies for the fast-track discovery of novel paramphistomicides,[34] antimalarial,[64, 65] and antitripanosomal/leishmania[5, 66, 67] compounds were as well conducted with this theoretical method.

On the other hand, other of our research teams has been studying the synthesis and reactivity of several families of heterocyclic betaines and salts. As a result of these and related studies, we have prepared many indazole,[68-73] indole,[74] cinnoline[75] and quinoxaline[76, 77] derivatives, several of which have shown interesting properties as trichomonacidal,[36, 61, 73, 78] antichagasic,[67, 73, 79] antimalarial[65] and antineoplastic[71-73] drugs.

Nowadays, the effort for the search of novel antiprotozoan drugs has increased considerably. However, existent effective broad spectrum antiparasitic agents? Therapeutics that are efficacious against most of species are interesting (and very important) because in the region of the world where these parasites are endemic do indeed overlap, and several infections are plausible and sometimes likely. We initially have been developed *"general"* (models are those based on activity datasets comprising diverse chemistries corresponding to a number of mechanisms of action[80]) QSAR models to description and prediction of the ***individual*** –antiprotozoan–infection.[5, 36, 60, 64, 65, 67, 78, 79, 81] Nonetheless, by using this approach a different model must be used to predict the *specific* antiparasitic activity for a given set of chemicals for every one of the antiprotozoan species. For this reason, is very important to develop a more *universal* model, which includes all chemicals reported as active against any protozoan parasite. This strategy will be permit us, to obtain *universal* models with a wide-broad application domain (antiprotozoan space) and maybe we also can to discovery drug-like agents with possible broad spectrum for their antiparasitic activity. Therapies that are able to treat several protozoan diseases would be practically attractive to person afflict by more one of parasite type or when the parasite involved is initially unknown.

In this report, we will explore the potential of ***TOMOCOMD-CARDD*** MDs to seek a QSARs-based ensemble classifier for antiprotozoan drug-like compounds from a

heterogeneous series of compounds. In the first step, we selected for the first time a wide-spectrum database of antiprotozoan drugs, which include compounds active against all kind of parasite protozoa *subphyla* and present diverse action modes. Next, the aforementioned MDs (specifically, the total and local non-stochastic and stochastic linear indices) were calculated for this large series of active/nonactive compounds and LDA was subsequently used to fit every individual classification function. Later, we developed a multi-agent QSAR classification system (ensemble classifier), in which the individual QSAR outputs are the inputs of the aforementioned fusion approach. Finally, the fusion model was used for the identification of a novel generation of lead-like antiprotozoans by using ligand-based virtual screening (LBVS) of small-molecules 'available' (with synthetic feasibility) in our 'in-house' library. A new molecular *sub-system* was then theoretically selected like promising lead series, which were subsequently synthesized, structurally characterized, and experimentally assayed. Here, we also describe the original synthesis and spectroscopic characterization of 10 molecules (new quinoxalinones) that had not been previously reported. The *in vitro* screening carried out here was design taking into account a battery of assays what include the most representative two different type of *subphylum* of protozoa parasites: 1) mastigophora (flagellata) and 2) apicomplexa (sporozoa). These "cell-based" (in this case parasite-based) assays suitable for describe a rather *complete profile* of antiprotozoan activity of these new chemicals.

Results and discussion

In silico Studies.

Here we will show three different computational experiments developed in this study. First we comments the result obtained in the construction of classification models

and their assembling like by using a fusion approach (multiagent-system). Each individual model was evaluated based on the guidelines set up in the Organization for Economic Cooperation and Development (OECD) principles.[82] They are intended to give some guidance and increase consistency in development and validation of (Q)SARs in order to be used for regulatory purposes. According to the OECD principles, a (Q)SAR should be associated with five points: (1) a defined endpoint, (2) an unambiguous algorithm, (3) a defined domain of applicability, (4) appropriate measures of goodness of fit, robustness and predictivity and (5) a mechanistic interpretation, if possible. This OECD principle form the basis of a conceptual framework for characterizing (Q)SARs, which assure that all necessary information is included and to describe the model characteristics in a transparent manner. Later, we describe the selection of new leads by using LBVS as well as the preparation of these new chemicals for simple and efficient methods of synthesis. Finally, the biological caracterization against four different species of protozoa parasites will be present in order to close the lead discovery cycle (experimental corroboration).

***Discussion on the Classification-based* Universal *QSAR for the Description of Antiprotozoan Activity*.** The development of discriminant functions that allows the classification of organic-chemical drugs as active or inactive is the key step in the present approach for the discovery of new wide-spectrum antiprotozoan agents. It was therefore necessary to select a training data set of active and inactive compounds containing broad structural variability and action modes as well as therapeutic uses. Therefore, the endpoint (*first principle*) here is the classification of chemicals into two different *experimental* classes: antiprotozoan (1) and non-antiprotozoan (-1) drug-like compounds. That is, antiprotozoan activity (drugs active against every species of

subphylum protozoa) is our define QSAR "endpoint," which can be measure and therefore modelled.

It is well-know that the general performance and extrapolation power of the learning methods decisively depends on the selection of compounds for the training series used to build the classifier model.[83] For this reason, and with the purpose of guarantee the molecular and pharmacological diversity we have selected a *benchmark* dataset composed by a great number of molecular entities, some of them reported as antiprotozoan[84-86] and the rest with a series of other pharmacological uses.[84, 85] We consider a large database of 680 drugs having great structural variability; 254 of them are active (antiprotozoan agents) and the others are non-antiprotozoan (426 compounds having other clinical uses, such as antivirals, sedative/hypnotics, diuretics, anticonvulsivants, haemostatics, oral hypoglycemics, antihypertensives, antihelminthics, anticancer compounds and so on). The classification of these compounds as "inactive" (without antiprotozoan activity) does not guarantee that any of these compounds present any antiparasitic activity no detected yet. The great structural variability of the selected training data set makes it possible, not only the discovery of lead compounds with determined mechanisms of antiprotozoan activity, but also with novel modes of action. It will be well-illustrated in this paper more below when we describe the third OECD principle (application domain).

Initially, two *k*-means cluster analyses (*k*-MCA) were performed for active and inactive series of chemicals, which permitted splitting the dataset (426 chemicals) into training (learning) and predicting (test) series.[87, 88] All cases were processed by using *k*-MCA in order to design training and predicting data series in a "rational" way. The main idea consists of carrying out a partition of either active or inactive series of chemicals in several statistically representative classes of chemicals. Thence, one may

select from the members of all these classes of training and predicting series. This procedure ensures that any chemical class (as determined by the clusters derived from *k*-MCA) will be represented in both series of compounds. Then, selection of the training and prediction sets was performed by taking, in a *random* way, compounds belonging to each cluster. The training set was composed by 204 antiprotozoans and 300 inactives from a set of 680 chemicals (504, ~75%). The resting group composed of 50 actives and 126 compounds with different biological activities was prepared as test data set for the validation of the models. These 176 (~25%) drugs were never used in the development of the classification models.

According to OECD Validation *Principle 2*, a (Q)SAR should be expressed in the form of an unambiguous algorithm. The intent of this principle is to ensure transparency in the description of the model algorithm. In this sense, in developing a method for predicting antiprotozoan activity, the first problem we face is how to represent the sample of a molecule. Here we used a defined mathematical algorithm, which is characterized in this case by two atom-based ***TOMOCOMD-CARDD*** MDs families (non-stochastic [$^{AP}f_k(\bar{x})$] and stochastic [$^{APs}f_k(\bar{x})$] linear indices).[32, 58, 63, 81, 89] This linear maps use a complete atomic properties (AP) scheme, which characterizes a specific aspect of the atomic structure (and k mean *order*, k = 1-15). The weights (atomic-labels) used in this work are those previously proposed for the calculation of the DRAGON descriptors,[90] i.e., atomic mass (AP = M), atomic polarizability (AP = P), atomic Mullinken electronegativity (AP = K) plus the van der Waals atomic volume (AP = V). All indices were also calculated taken into account all H-atoms in the molecule, i. e., $^{AP}f_k^{H}(\bar{x})$ and $^{APs}f_k^{H}(\bar{x})$ for non-stochastic linear indices and their stochastic counterpart, respectively. Two local (L) atom-type indices for heteroatoms

(group = heteroatoms (E): E = S, N, O), not considering [$^{AP}f_{kL}(\bar{x}_E)$] and considering [$^{AP}f_{kL}(\bar{x}_E)$] H-atoms in the molecule, were computed too.

The representative selection of training set permit continues to the next step, the finding of the classification functions to discriminate between active and inactive. For this we select the LDA as statistical technique due to it's broadly use and simplicity. As we describe above, LDA also is a statistical technique with a define algorithm, therefore on the OECD basis the *second principle* is proposed as being satisfactorily met.

All Classification-based QSAR equations derived by using forward stepwise LDA and all set of total and local atom-based linear indices computed are shown below:

$$Class = -3,84 - 3,14 *10^{-4}\, {}^{M}f_5^{H}(\bar{x}) + 2,79*10^{-2}\, {}^{M}f_1(\bar{x}) + 4,19*10^{-3}\, {}^{M}f_2(\bar{x}) + 2,72*10^{-8}\, {}^{M}f_{12}(\bar{x}) - 2,45*10^{-3}\, {}^{M}f_{4L}{}^{H}(\bar{x}_E) + 4,23*10^{-6}\, {}^{M}f_{10L}{}^{H}(\bar{x}_E) - 2,40*10^{-8}\, {}^{M}f_{14L}(\bar{x}_E) \quad (1)$$

$$Class = -3,97 - 2,32 *10^{-5}\, {}^{P}f_8^{H}(\bar{x}) + 6,23*10^{-3}\, {}^{P}f_5(\bar{x}) - 1,87*10^{-4}\, {}^{P}f_9(\bar{x}) + 6,47*10^{-6}\, {}^{P}f_{12}(\bar{x}) - 6,55*10^{-8}\, {}^{P}f_{15}(\bar{x}) + 2,37*10^{-6}\, {}^{P}f_{11L}{}^{H}(\bar{x}_E) - 1,46*10^{-8}\, {}^{P}f_{15L}(\bar{x}_E) \quad (2)$$

$$Class = - 4,03 - 1,34 *10^{-9}\, {}^{V}f_{14}^{H}(\bar{x}) + 3,37*10^{-3}\, {}^{V}f_1(\bar{x}) + 8,23*10^{-9}\, {}^{V}f_{13}(\bar{x}) - 2,47*10^{-3}\, {}^{V}f_{4L}{}^{H}(\bar{x}_E) + 1,78*10^{-7}\, {}^{V}f_{12L}{}^{H}(\bar{x}_E) + 1,84*10^{-2}\, {}^{V}f_{2L}(\bar{x}_E) - 4,12*10^{-9}\, {}^{V}f_{15L}(\bar{x}_E) \quad (3)$$

$$Class = - 3,84 - 1,36 *10^{-4}\, {}^{K}f_8^{H}(\bar{x}) + 3,42*10^{-5}\, {}^{K}f_9^{H}(\bar{x}) + 0,27\, {}^{K}f_0(\bar{x}) - 6,76*10^{-3}\, {}^{K}f_3(\bar{x}) - 6,96*10^{-2}\, {}^{K}f_{2L}{}^{H}(\bar{x}_E) + 3,76*10^{-5}\, {}^{K}f_{9L}{}^{H}(\bar{x}_E) - 1,71*10^{-8}\, {}^{K}f_{15L}(\bar{x}_E) \quad (4)$$

$$Class = - 4,06 + 2,8 *10^{-8}\, {}^{M}f_{12}(\bar{x}) - 4,53*10^{-8}\, {}^{P}f_{15L}(\bar{x}_E) + 1,34*10^{-7}\, {}^{V}f_{12L}{}^{H}(\bar{x}_E) + 9,23*10^{-3}\, {}^{V}f_{2L}(\bar{x}_E) - 1,36*10^{-5}\, {}^{K}f_8^{H}(\bar{x}) + 0,14\, {}^{K}f_0(\bar{x}) - 6,35*10^{-2}\, {}^{K}f_{2L}{}^{H}(\bar{x}_E) \quad (5)$$

$$Class = -3,13 - 5,28 *10^{-2}\, {}^{Ms}f_2^{H}(\bar{x}) + 0,26\, {}^{Ms}f_2(\bar{x}) - 0,18\, {}^{Ms}f_{10}(\bar{x}) + 0,10\, {}^{Ms}f_{1L}{}^{H}(\bar{x}_E)$$

$$- 5{,}46*10^{-2}\ {}^{Ms}f_{1L}(\bar{x}_E) - 0{,}20\ {}^{Ms}f_{2L}{}^{H}(\bar{x}_E) + 0{,}15\ {}^{Ms}f_{14L}(\bar{x}_E) + 3{,}73\ {}^{Ms}f_{3L}(\bar{x}_{H-E}) \quad (\mathbf{6})$$

$$Class = -4{,}00 + 0{,}74\ {}^{Ps}f_{1}{}^{H}(\bar{x}) - 0{,}72\ {}^{Ps}f_{2}{}^{H}(\bar{x}) - 0{,}56\ {}^{Ps}f_{11}{}^{H}(\bar{x}) + 0{,}87\ {}^{Ps}f_{4}(\bar{x})$$
$$-2{,}12\ {}^{Ps}f_{1L}{}^{H}(\bar{x}_E) + 1{,}12\ {}^{Ps}f_{2L}{}^{H}(\bar{x}_E) + 1{,}31\ {}^{Ps}f_{3L}{}^{H}(\bar{x}_E) \quad (\mathbf{7})$$

$$Class = -3{,}79 + 0{,}14\ {}^{Vs}f_{0}{}^{H}(\bar{x}) - 0{,}08\ {}^{Vs}f_{2}{}^{H}(\bar{x}) - 0{,}03\ {}^{Vs}f_{7}{}^{H}(\bar{x}) - 0{,}73\ {}^{Vs}f_{3}{}^{H}(\bar{x})$$
$$+ 1{,}94\ {}^{Vs}f_{5L}{}^{H}(\bar{x}_E) + 0{,}16\ {}^{Vs}f_{6L}{}^{H}(\bar{x}_E) - 1{,}30\ {}^{Vs}f_{7L}{}^{H}(\bar{x}_E) - 0{,}07\ {}^{Vs}f_{0L}(\bar{x}_E) \quad (\mathbf{8})$$

$$Class = -4{,}27 + 3{,}25\ {}^{Ks}f_{5}{}^{H}(\bar{x}) - 3{,}23\ {}^{Ks}f_{7}{}^{H}(\bar{x}) + 2{,}97\ {}^{Ks}f_{1L}{}^{H}(\bar{x}_E) + 3{,}34\ {}^{Ks}f_{6L}{}^{H}(\bar{x}_E)$$
$$- 1{,}62\ {}^{Ks}f_{1L}(\bar{x}_E) - 4{,}29\ {}^{Ks}f_{6L}(\bar{x}_E) - 13{,}81\ {}^{Ks}f_{6L}{}^{H}(\bar{x}_{E-H}) + 60{,}01\ {}^{Ks}f_{10L}{}^{H}(\bar{x}_{E-H})$$
$$- 46{,}26\ {}^{Ks}f_{12L}{}^{H}(\bar{x}_{E-H}) \quad (\mathbf{9})$$

$$Class = -3{,}93 - 9{,}25*10^{-2}\ {}^{Ms}f_{2L}(\bar{x}_E) - 0{,}98\ {}^{Ps}f_{1L}{}^{H}(\bar{x}_E) + 0{,}18\ {}^{Vs}f_{6L}{}^{H}(\bar{x}_E)$$
$$+ 5{,}98*10^{-2}\ {}^{Vs}f_{0L}(\bar{x}_E) + 10{,}99\ {}^{Ks}f_{10L}{}^{H}(\bar{x}_{E-H}) - 11{,}23\ {}^{Ks}f_{12L}{}^{H}(\bar{x}_{E-H}) \quad (\mathbf{10})$$

$$Class = -3.97 - 1{,}89*10^{-3}\ {}^{M}f_{4L}{}^{H}(\bar{x}_E) + 2{,}59*10^{-2}\ {}^{V}f_{2L}(\bar{x}_E) - 6{,}96*10^{-2}\ {}^{K}f_{2L}{}^{H}(\bar{x}_E)$$
$$+9{,}24*10^{-6}\ {}^{K}f_{9L}{}^{H}(\bar{x}_E) - 0{,}72\ {}^{Ps}f_{1L}{}^{H}(\bar{x}_E) + 0{,}10\ {}^{Vs}f_{7L}{}^{H}(\bar{x}_E) \quad (\mathbf{11})$$

In total were obtained eleven models, the first four equations (1-4) developed with the non-stochastic bond-based linear indices and the other four first four (6-9) perform with the stochastic MDs. Overall performances of all the obtained models are given in Table 1, together with the Wilks' statistics (λ), the square of the Mahalanobis distances (D^2), and the Fisher ratio (F). The models selected show to be statistically significant at *p*-level< 0.001. This Table also shown the obtained result for the equations **5** and **10** of the last five models in both cases (non-stochastic and stochastic molecular fingerprints) resulting in a combination of all pairs of atom weights (atomic labels). In addition, the equation **11** was carried out by using all set of MDs (mixing non-stochastic and stochastic linear indices) and was the best models in learning set (see Table 1).

Table 1 comes about here (see end of the document)

The fitted models **5** and **10**, resulting of the combination of weighting schemes for the non-stochastic and stochastic atom-level linear indices, respectively, as well as the equation **11** (mixing non-stochastic and stochastic indices) exhibit the best results, how can be observed in Table 1. These best two equations based on both individual set of linear indices (Eqs. **5** and **10**) correctly classified the 91.27% of the training set, and showed values of the Matthews correlation coefficients (***C***) of 0.82. However, equation **5** (non-stochastic linear indices) showed more **false positive rate** than equation **10**, fitted by using only stochastic MDs. However, the best result is performed when all set of MDs was used. The equation **11** showed 93.06% of global good classification and a ***C*** of 0.86. The most common parameters in medical statistics for all the models are depicted in the same Table 1. The classifications of every compound in learning series are shown in Table SI**1** of Supporting Information. Likewise a plot of the ΔP% (see Experimental Section) for the entire training set using the best models **11**, is illustrates in Figures 1.

Figure 1 comes about here (see end of the document)

Other crucial problem in chemometric and QSAR studies is the definition of the Applicability Domain (AD) of a classification or regression model. "*Not even a robust, significant, and validated QSAR model can be expected to reliably predict the modelled property for the entire universe of chemicals. In fact, only the predictions for chemicals falling within this domain can be considered reliable and not model extrapolations*".[91] Therefore, the next step of this report was developed a study to access to chemical's scope of our models (*principle 3: Defined Domain of Applicability*). The AD is a theoretical region in chemical space, defined by the model descriptors and modelled response, and thus by the nature of the chemicals in the training set, as represented in

each model by specific MDs. That is to say, AD of the QSAR model is "the range within which it *tolerates* a new molecule".[92]

For RLM and ADL, a multiple predictor problems with normally distributed data, the distance-based measures, like ***leverage*** (*h*) is one of most used (see Experimental Section).[93, 94] The warning leverage, *h**, is a *critical value* or *cut-off* to consider the prediction made for the model for a specific compounds in dataset. To visualize the AD of a QSAR model, a double ordinate Cartesian plot of cross-validated residuals (first ordinate), standard residuals (second ordinate), and leverages (Hat diagonal: abscissa) values (*h*) defined the domain of applicability of the model as a squared area within ±3 band for residuals and a leverage threshold of $h^* = 0.042$ for antiprotozoan activity (i. e., Eq. **11**). This plot, so-called Williams scheme can be used for an immediate and simple graphical detection of both the response outliers (i.e., compounds with standardized residuals greater than three standard deviation units, >3σ) and structurally influential chemicals in a model ($h>h^*$). For instance, Figure 2 shows the Williams plot of Eq. **11** as a simple example. As can be noted in Figure 2, almost all chemicals used lie within this area. Actually, some chemicals like in test set, Trypan red ($h = 0.371$) and Dithiophos ($h = 0.156$) have leverage very higher than the threshold but show residuals within the limits. These active and inactive compounds are outside of application domain of this model and these chemicals can influence model parameters. Considering this fact, we must check the effect of withdrawal of these compounds on the model performance. When we study the new parameters of the model after removal of these chemicals we detected no significant variation as well as the model performance. Therefore, the influence of these compounds in not critical neither for model parameters nor performance. Consecuently, their removal in not justified. In addition, Sch 18545 (antiprotozoan with *h* of 0.113) and Siccamid (nonantiprotozoan with *h* of 0.109) had

the *h* most high in training set. However, these compounds presented residuals rather low than the previous ones, and how these chemicals are in the same experimental space (inside of this range) that others 20 cases in training set (slightly exceed the critical h^* but are very close to other chemicals of the training set and in this same zone), which slightly exceed the critical hat value (vertical line), slightly influential in the model development: the predictions for new compounds in this sense situation (for instance, included in a external test set, where there are 13 cases that slightly exceed the critical h^* value) can be considered as reliable as those of the training chemicals and the possible erroneous prediction could probably be attributed to wrong experimental data rather than to molecular structure. Finally, two compounds Myralact (σ =3.09) and Tosulur sodium (σ = 3.187), which are cases of training and test sets, depicted outlier behibour with standardized residuals greater than three standard deviation units. That is to say, both chemicals was wrongly predicted (>3σ); it is these two compounds as well as the inicially two compounds (Trypan red and Dithiophos) are completely outside the AD of the model, as defined by the Hat vertical line (high *h* leverage value). Thus, four compounds that are either a response outlier or a high leverage chemical. In closing, the model can be used with high accuracy in this applicability domain.[91, 94] In the next section we re-taken this analysis in order to determine the reability of prediction for molecules selectioned like rather good candidates in *virtual* screening protocols.

Figure 2 comes about here (see end of the document)

The model validation (*Principle 4: Statistical Validation*) is other key features in good QSAR practice regarding with diagnostic of developed models. In this sense, a QSAR model should be associated with an appropriate measures of goodness-of-fit, robustness and predictivity.[87, 95, 96 97, 98] Both first they are considered as *internal validation*, while the later is considered as *external validation*. The evaluation of

performance of models by using external validation (one o more external test sets) can be considered as a *superior alternative* because the good behaviour of models in internal experiments are the necessary but not sufficient condition for the model to have high predictive power. That is, the predictivity can be claimed only if the model successfully applied to prediction of the external test series chemicals, which were not used in the model development. For this reason, in this report we only describe the external performance evaluation by using a prediction set of active and inactive compounds.

Table 2 comes about here (see end of the document)

The key parameters for statistical diagnostic of all obtained models are present in Table 2. As can be observe, the prediction performance for LDA-based QSAR models in the test set was adecuate. Here, the results shown that the equations obtained with non-stochastic indices are better than models derived with stochastic MDs. In addition, the best LDA-based QSAR is the equation **11,** with a accurancy of 92.05% *vs* 85.80% depicted by models **5**. Finally, the classifications of every compound in prediction series are illustrate as Supporting Information (Table SI**2**). Likewise a plot of the ΔP% (see Experimental Section) for the entire test set by using the best models **11**, is show in Figures 3.

Figure 3 comes about here (see end of the document)

Therefore, the performance of our computational approach was assessed on the basis on sound design. The outcomes were carefully evaluated in light of classification parameter and the behaviour was rather good. That is, the obtained acceptable values validate the models for their use in LBVS, taking into account the acceptable values above 75% in all the test set, which is considered as a suitable threshold limit for this kind of analyses. Hence OECD *Validation Principle 4* is fully met. The last *principle 5*

(*Mechanitic Relevance*, if it possible) is rather difficult to address in this report, due to the nature of database used for develop of QSARs.

Drug(Lead)-like Discovery by Virtual (*In Silico*) Screening and *Dry* Selection: *To be or not to be.* The ligand-based methods are supported in the principle of similarity –similar compounds are assumed to produce similar effects[99]– and serve to model the complex phenomena of molecular recognition. Similarity-based methods are cornerstones of chemoinformatic and computer-aided pharmaceutical research. To this effect, LBVS has been used to identify novel active compounds in many biological applications. This indicates that 'similarity' methods should have substantial 'selectivity' in recognizing diverse active compounds.[100] Current purposes to integrate chemoinformatics into "real-life" applications, to step-ahead in drug discovery are of main importance nowadays. Following this aim, and because drug discovery is a complex phenomenon that requires the evaluation of large amounts of chemical data, it could be said that *in silico* predictions are suitable to detect the biological activity under study.

The algorithm described above, and the obtained good results prompted us to make *in silico* evaluations of all the chemicals contained in our '*in-house'* collections of indazole, indazolols, indole, cinnoline, and quinoxaline derivatives (as well as other new related chemicals and their derivatives), which have been recently obtained by our chemical synthesis team. On the basis of computer-aided predictions we selected potential antiprotoazoan leads (*virtual hits*). The following criteria were used for the hits' selection: 1) compounds were selected as *hits* if the value of posterior probability of possessing antiprotozoan activity exceeded 15% ($\Delta P > 15\%$) by all LDA-based QSAR models (fusion approach or multi-classification system), and 2) If, among the compounds designed (or that it will obtain in our laboratory) by our chemical team, too

many similar compounds satisfied criterion 1, then only several representative structures were selected.

Here, we perform *in silico* mining of our library and some heterocyclic leads were identified (selected) like novel antiprotozoan by using the discriminant functions obtained through the ***TOMOCOMD-CARDD*** method and LDA data-mining technique as an ensemble classifier, $\boldsymbol{C_E}$. That is, here every individual classifier ($\boldsymbol{C_I}$) is fused into the $\boldsymbol{C_E}$ through a voting system, where the individual output of $\boldsymbol{C_I}$ are used like input of $\boldsymbol{C_E}$, which will have a voting *score* for the query molecules **M** (for more detail see Experimental Section). To provide an intuitive picture, a flowchart to show how these $\boldsymbol{C_I}$ are fused into the $\boldsymbol{C_E}$ is given in Figure 4.

Figure 4 comes about here (see end of the document)

One series of compounds (quinoxalinones derivatives) was selected as antiprotozoan lead-like compounds, showing a good agreement between the *in silico* predictions and *in vitro* assays in several cell(parasite)-based tests (see more below). The values of ΔP% for this subset are depicted in Table 3.

Table 3 comes about here (see end of the document)

This result shows an experimental example of QSAR application for the development of drug discovery; besides, it could be an effective help for further design and optimization in this type of lead compounds as a way to improve the antiprotozoan activity, from the selection of *hits*, followed by the elucidation of the behaviour in the pharmacological and toxicological assays.

However, it is generally acknowledged that QSARs are valid only within the same domain for which they were developed. In fact, even if the models are developed on the same chemicals, the AD for new chemicals can differ from model to model, depending on the specific MDs. One of the main aims of the present work was to develop a model

for predicting antiprotozoan activity at early stages of drug discovery and development. Consequently, one may not pretend *to extrapolate* the use of these models to other kind of class-antiprotozoan making uncertain predictions in conditions very different to those fixed to derive the model.[93, 94] Therefore, the chemical designed in this studies only were syntetized and posterior *in vitro* evaluated after that they were ploted into the AD of obtained models. For instance, another William plot (Figure 5) of Eq. **11** (with the training set and quinoxalinone series discovered as novel antiprotozoan leads was carried out) as a simple example. As can be noted in Figure 5, all quinoxalinones used lie within this area, which ensures great reliability for the prediction of this kind of leads used in the virtual screening. That is to say, all new leads fall within the applicability domain of the model and so the predictions are reliable.

Figure 5 comes about here (see end of the document)

This proves the good assessment for the classification of these quinoxalinones as novel antiprotozoan leads. Therefore, this model can be used high accuracy for new compound predictions in this applicability domain.[93, 94]

Chemistry Result.

Owing to their direct involvement with the present paper, special mention deserves our study on the synthesis and biological activity of a series of 3-alkoxy-1-[5-(dialkylamino)alkyl]-5-nitroindazoles,[73] as well as a previous work on the synthesis and reactivity of quinoxalinium salts prepared from substituted acetanilides through intramolecular quaternization reactions.[76]

On these bases and taken into consideration the early *in silico* selection of quinoxaline *molecular sub-system* like promisorial antiprotozoan lead series, we decided the preparation (syntesis and spectroscopical caracterization) and futher biological efficacy of 7-nitroquinoxalin-2-ones **9-18**, carrying at position 4 a 5-

(dialkylamino)pentyl chain similar to that of the mentioned indazole derivatives, according to the synthetic pathway shown in the Scheme as well as the spectroscopical characterization of these compounds and intermediates, and the further study of their biological efficacy.

Scheme comes about here (see end of the document)

Thus, treatment of substituted aniline **1** with bromoacetyl bromide afforded 2-bromoacetanilide **2**, which cyclized easily to the spiro quinoxalinium bromide **5**. This salt, as well as the corresponding 1-methyl analogue **6**, could also be prepared by treatment of the previously prepared[76] chlorides **3** and **4** with hydrobromic acid through a halogen exchange reaction. Piperidine ring of salts **5** and **6** was then cleaved in refluxing nitromethane to yield the corresponding 4-(5-bromopentyl)quinoxalinones **7** and **8**.

Finally, treatment of compounds **7** and **8** with the required secondary amines (dimethylamine, pyrrolidine, piperidine, homopiperidine or 1,2,3,4-tetrahydroisoquinoline) afforded the final 4-[5-(dialkylamino)pentyl]-7-nitroquinoxalin-2-ones **9**-**18**, which were isolated as the corresponding hydrobromides. The previously prepared[76] chloro analogues of **7** and **8** were rather unreactive under the conditions used in this work (see **Experimental Section**) and were not appropriate for the preparation of the desired final compounds.

The structure of all compounds has been established on the basis of their analytical and spectral data. The latter are similar to those of related 1-[5-(dialkylamino)alkyl]indazoles,[73] quinoxalines and intermediates[76] previously prepared by us. Thus, NMR spectra of 2-bromoacetanilide **2** show that this compound, like the corresponding chloro analogue,[76] appears in $CDCl_3$ solution as the *Z*-rotamer. On the other hand, owing to the rigidity of spiro bromides **5** and **6**, NCH_2 protons of piperidine

rings are anisochronic and, according to their different coupling patterns, they can be distinguished as equatorial (H_e) and axial (H_a). Similar features were observed for the cyclic secondary amine-derived final products **10-13** and **15-18**, accordingly to their structure of tertiary ammonium bromides. NCH_2 protons of piperidine rings of compounds **11** and **16** can also be distinguished as H_a and H_e. Nevertheless, the assignment (equatorial or axial) of other protons of piperidine rings and protons of pyrrolidine (**10**, **15**), homopiperidine (**12**, **17**) and 1,2,3,4-tetrahydroisoquinoline (**13**, **18**) derivatives is not easy; when separate signals are observed, they have been mentioned in the description of ^{1}H NMR spectra as H_A and H_B.

In Vitro *Screenning and* Wet *Evaluation.*

In the present section we describe the main results obtained in the experimental assays (*wet evaluation*) in four different protozoan-parasite tests of the new chemicals selected like lead series in our *in silico* experiment. Here, we developed a *wet* screening taking into account a battery of tests, that include the most representative two different type of *subphylum* of protozoa parasite: 1) *T. vaginalis and T. cruzi,* which belong to mastigophora (flagellata) *subphylum* and also, 2) two different apicomplexa (sporozoa) parasites: *P. falciparum* and *T. gondii*. These parasite-based tests will permit to depict a rather *complete profile* of antiprotozoan activity of these new compounds.

Firstly, we evaluate the designed compounds against *T. vaginalis* and *T. cruzi.* In the case of the later parasite, the epimastigote form was used in the *in vitro* experiment taken into consideration that this form is an obligate mammalian intracellular stage.[101] In addition, unspecific cytotoxicity to macrophages were tested for all compounds. The *in vitro* efficacy against *T. vaginalis* and *T. cruzi* (as well as unspecific cytotoxicity) are shown in Table 4 and 5, respectively.

Table 4 and 5 comes about here (see end of the document)

The specific activity against *T. cruzi* and *T. vaginalis* are expressed as percentages of anti-epimastigote activity and growth inhibition (cytostatic activity), respectively. Cytocidal activity (percentage of reduction with respect to the control) against *T. vaginalis* is shown in brackets. Metronidazole and Nifurtimox were used as trichomonacidal and trypanocidal reference drugs, correspondingly. Unspecific cytotoxic activity to macrophages is expressed as cytotoxicity percentage.

In general, all chemicals showed low unspecific cytotoxicity, except for compounds **13**, **17,** and **18** at 100 μg/mL. Most of the compounds tested, exhibited a trichomonacidals activity near to 100% (**11-18, 14**) at the higher concentration assayed (100 μg/mL). Only compound **10** and **9** were inactive at this level. However, only chemicals **15-17** showed cytocidal activity against *T. vaginalis* at 10 μg/mL after 24 h of contact. These derivatives showed rather good antiprotozoan action at this level (near 90%; percentage of reduction with respect to the control), but this effect does not appear at 48 h of contact. At this time, only at the first concentration of 100 μg/mL **11-18** were actives.

In the same form, most of the tested compounds also exhibited a trypanocidal activity of 80 to 100% (**10-13** and **16**) at 100 μg/mL. This activity is not unspecific, since all of them, except for compound **13**, showed cytotoxicity lower than anti-epimastigote activity (see Table 5). However, the trypanocidal activity dramatically decreases at the lower dose. Only compound **16** retained a 60% of activity at 10 μg/mL; at this concentration no unspecific cytotoxicity was shown for this compound.

Comparing the activity against the two species of parasites (as well as cell toxicity) of these ten compounds it is possible to conclude that **15-17** are the best chemicals. Specifically, **16** was the most active compound in both parasite and therefore, this chemical can be taken as *hit* for anti-mastigophora *subphylum* parasites.

From experiments, we can to do some relevant conclusions about structure-activity relationship. For instance, the methyl group at N-1 (**14-18**) enhances the activity against both species of flagellate protozoan parasites. The 6-member ring in substituent at N-4 (**11** and **16**) is the best chemical function, and to open this ring in lethal for bioactivity (**9** and **14**) as well as the use of tetrahydroisoquinoline moiety (**13** and **18**), which also raise the toxicity of this lead series (see last column in Table 5).

In the second step, we evalued the same compounds against the more human's important protozoan *subphylum*. Here, we initially tested the efficacy of these chemicals against tachyzoites form of *Toxoplasma gondii* (RH strain).[102, 103] This overall result archived in this experiment is depicte in Table 6.

Table 6 comes about here (see end of the document)

Compounds **10-12** show toxoplasmicidal effects at concentrations of 1 mM and 500 μM. Compound **17** was active against the parasite at 1mM concentration. The evaluation of the parasites by light microscopy (data not shown) demonstrated that the four drug-like compounds seem to protrude the organelles of the thachyzoites. The damage to the tachyzoites with compound **11** was more aggressive than the ones caused by the others three compounds. The assays with the evaluated compounds and controls were made in triplicate. Negative controls had 96% viability. Compounds **13** and **18** were not evaluated because their dilution in MEM causes precipitation.

Under the conditions that this assay was made, we concluded that some of the tested compounds seem to have activity against *Toxoplasma gondii* purified tachyzoites. It was found that **11** had the most potent anti-toxoplasma activity at high concentrations. These results suggested that the compounds **10-12** of this series may be chosen as possible candidates in the development of toxoplasmicidal chemotherapy. More studies need to be done to evaluate the effect of the chemicals on the structural, functional and

virulent properties of *Toxoplasma gondii in vitro* and *in vivo* in order to design new drugs against these reemerging parasitic zoonoses. These studies are being carried out and will publish in forthcoming paper. In conclusion, the compound **11**, with the same function at N-4 that **16**, but with an H-atom in N-1 was the most active compounds. This result indique that H-atom in the N-1 is necessary for anti-toxoplasma activity in opposition to obtained for *flagellate* parasites, where the methylation of this N-atom was desire. Maybe, it is a logical result if we taken into account that these parasites belong to two different protozoan *subphylum*.

Finally, these compounds were assayed in two different tests for antimalarial screening. The first techniques used was a cell- and enzyme-free *in vitro* assay, the *so*-called: ferriprotoporphyrin IX biocrystallization inhibition test (FBIT).[104] During their digestion of host cell haemoglobin, intraerythrocytic malaria parasites produce large amounts of toxic ferriprotoporphyrin IX (FP). The inhibition of biomineralisation of FP to β-hematin by some antimalarial compounds such as chloroquine underlies their action mode and in this sense, it can be used to give a criterion of potential antimalarial character.[65, 104] The global results for the selected chemicals in this enzymatic *in vitro* model are depicted in Table 7.

Table 7 comes about here (see end of the document)

From ten compounds, only 3 cases (**13, 17** and **18**) showed IC_{50} values lower than 2.0 μg/mL, resulting actives in the biomineralisation microassay. The remaining seven, resulted inactive ones. In this assay, any compound resulted more active than chloroquine (see Table 7). The order according to activity is **18 >13 >17.** However, these chemicals had unspecific cytotoxicity at 100 μg/mL.

Afterwards, a cell-based approach was also used to evaluate the *in vitro* efectivity of the designed series. This second *in vitro* cell-based assay was carried out by using a

radioisotopic microtest in *Plasmodium faciparum* (strain 3D7).[105] Here, every compound was evaluated against cultured intraerythrocytic asexual forms of the human malaria parasite *P. falciparum*. The uptake of [G-^{3}H]hypoxanthine by parasitized erythrocytes in the microtiter plates was used as an indicator of drug activity. As can be seen in Table 7, compound **18** also was active in this *wet* evaluation, while those chemicals **13** and **17** were inactives. However, compounds **12** showed rather activity in this cell assay. This compound had low cytotocicity and was also activity againt *T. gondii,* therefore this chemical core and SAR result (H atom at N-1 and a 6-membered ring at N-4) can be considered as an important starting point to the design of novel antiapicomplexa drugs. In this sense, new refining algorithms are needed for optimizing the farmacological, toxicological and physico-chemical properties.

In summary, these results can be considered as a promising starting point for the future design and refinement of novel compounds with higher anti-protozoan activities and low toxicity. Although compounds **15-17** (lead series for anti-mastigophora *subphylum*) and **10-12** (lead serie for anti-apicomplexa *subphylum*) were active at higher doses than their respective reference drugs. Analysing all these *in vitro* results, it is clear to see that further refinement algorithms are needed to identify the ways in which the activity and ADMETox of the present chemical core can be optimized. Therefore, these chemicals, manly **11** (**12**) and **16**, can be taken as *hits*, which are amenable for further chemistry optimization in order to derive the appropriate combination of potency, pharmacokinetic properties, toxicity etc., as well as good activity in animal models.

Conclusion

The integration (aligning) of *dry* and *wet* screening for diverse compounds libraries is an essential part of the antiprotozoan lead discovery effort. The results of our *in silico* prediction and posterior *in vitro* screening by using a battery of parasites-cell assays are encouraging and show that progress may be made throught this kind of approach. Within this one set of *in house* library, we have identified 10 novel chemicals not yet reported (*virtual hits*) like antiprotozoan lead. All novel quinoxalinones were them synthesized employing simple and efficient methods of preparations. The spectral (structural) caracterization is also presentated in this report. Finally, the biological evaluation shows that most of the compounds tested, exhibited adecuate antiprotozoan activities against four different kinds of parasites (*T. vaginalis, T. cruzi, T. gondii* and *P. falciparum*). In general, all chemicals showed low unspecific cytotoxicity, except for compounds **13**, **17**, and **18** at 100 μg/mL. However, the most active compound, **11(12)** and **16**, do not present cytotocicity in macrophages cell at any level. This chemicals show preliminary evidence of good and selective wide-range for antiprotozoan activities with potential for scaffold optimization.

A future perspective

The development of a new drug is a lengthy and complex process. The identification of an appropriate lead molecule is the most critical component of this phase. Over the past few decades, a primary source for novel leads has been the high-throughput screening (HTS) of compound libraries. The advent of virtual screening (ligand- or structure-based) methods to identify a reduced number of molecules with increased potential for bioactivity to be experimentally evaluated — has emerged both as a complementary and alternative method to HTS.

Taking into account that QSAR has been applied extensively in the recent years to find predictive models for activity of bioactive agents, and that researches in many biological activities are based on traditional *trial-error* methods until nowadays, we have performed a QSAR study to discriminate antiprotozoan compounds from inactive ones. The main idea on the integration of emerging QSAR research strategies is to identify new approaches to decide which molecular structures to synthesize and ultimately pursue in the drug-discovery setting. The impact on decreasing the likelihood of entering the lead-drug discovery processes is one of the topics to be assessed. To this effect, here we have shown how the combination of validated QSAR-modeling and LBVS, could be successfully used as innovative technologies, to ensure high expected *hit* rates in the discovery of new bioactive compounds. In future outlooks, these models which relate the chemical structure with a specific endpoint, could be programmed into expert systems helping in exhaustive search of bioactive molecules within huge chemical libraries. That is to say, the preliminary identification of novel antiprotozoan leads in this work is promising and strongly supports the LBVS of additional compounds libraries; manly of chemicals with diverse scaffolds is an important strategy to continue exploring. In fact, the assemble clasifier present here will be use to identifie new antiprotozoan from database of well-know drugs already approved for human use for potential 'off-label' antiparasitic efficacy. The logic of this approach is that *hits* from such screens are low-hanging fruit that will require less development before they are able to enter clinical trials as antiparasitics. Some work in this direction is now in progress and will be published in a forthcoming paper.

The mode action of the novel quinoxalinones described in this study is a question that has not been addressed. While this is beyond the scope of this report, it is extremely relevant, and we are currently following up on the top leads. Along these lines, question

of ADMETox are all avenues we will explore, as they will illuminate of future study on optimization of these leads, but first explored all these issues, in parallel way, by using theorethical models. In this sense, our research group is working in the application of new 3D MDs and data mining techniques to these problems. We are also interested in apply our old and new MDs to codify action modes of antiprotozoan chemicals. Also we have planed to concentrate our efforts in the use of more sophisticated statistical techniques to be used with the ***TOMOCOMD-CARRD*** MDs in order to describe the activity of organic compounds against important pharmacological *targets* of antiprotozoan drugs. That is, we will develop models to predit the biological response to specific antiprotozoan molecular targets and so, complete a computational system that permit the identification as well as optimization of new leads in parallel manner. Another direction to explore in the future study is the multi-optimization (approach) in order to caracterizing the biological response of one chemical versus multitarget, *for instance*: different species, different molecular targets, and so on. Continuation of the kind of library screening that we have presented here and the future analysis that we will carried out of proposed therapeutics are potentially effective strategies to help fight the worsening protozoan illness plight.

Experimental Section

Computational Strategies

Data set and classification strategy. A benchmark dataset usually consists of a learning (or training) dataset and an independent testing dataset. The learning dataset is one of the important components for a statistical predictor because it is used for training the predictor's ''engine,'' whereas the testing dataset is used for examining the predictor's accuracy via an external test.[106] The *benchmark* dataset was composed by

680 drugs having great structural variability; 254 of them are active (antiprotozoan agents) and 426 inactive compounds (drugs having other clinical uses).[84-86]

Representation of molecules samples. Several kinds of representations are generally used in this regard, all well-know like *molecular descriptor* (MDs) or *molecular indices*. This parameters are numbers that characterize a specific aspect of the molecule structure.[107] The *so*-called topological (and topo-chimical) indices are among the most useful MDs known nowadays.[108, 109] These theoretical indices are numbers that describe the structural information of molecules through graph theoretical invariants and can be considered as structure-explicit descriptors.[110]

In the present report, a novel 2D ***TOMOCOMD-CARDD*** MDs family, namely atom, atom-type, and total linear indices were used in order to codify the molecular structure of every molecule en dataset. These MDs are based on the calculation of linear maps (linear form) in $\Re^n$ in canonical basis sets.[32, 58, 63, 81, 89] The computation of the non-stochastic and stochastic linear indices is develop by using the k^{th} "nonstochastic and stochastic graph–theoretical electronic-density matrices" $\mathbf{M}^k$ and $\mathbf{S}^k$, correspondingly, as matrices of the mathematical forms.[32, 58, 63, 81, 89] These matricial operators are graph-theoretical electronic-structure models, like the ''extended Hückel MO model''. The $\mathbf{M}^1$ matrix considers all valence-bond electrons (σ- and π-networks) in one step, and their power k (k = 0, 1, 2, 3,...) can be considered as an interacting-electronic chemical-network in step k. The present approach is based on a simple model for the intramolecular (stochastic) movement of all outer-shell electrons. The teorethical scaffol of this atom-based MDs and their use to represent small-to-medium size organic chemicals as well as QSAR and drug design studies has been explained in some detail elsewhere.[32, 58, 63, 81, 89]

Computational Methods: *TOMOCOMD-CARDD* approach. *TOMOCOMD* is an interactive program for molecular design and bioinformatics research, developed upon the base of a user-friendly philosophy.[24] In this reports, we only used the *CARDD* (*C*omputed-*A*ided '*R*ational' *D*rug *D*esign) subprogram. All MDs [total and local (both atom and atom-type) non-stochastic and stochastic linear indices were calculated in this software.

Chemometric Studies. The statistical software package STATISTICA was used to develop the *k*-MCA.[111] The number of members in each cluster and the standard deviation of the variables in the cluster (kept as low as possible) were taken into account, to have an acceptable statistical quality of data partitions into the clusters. The values of the standard deviation between and within clusters, the respective Fisher ratio and their *p* level of significance, were also examined.[88]

LDA was also carried out with the STATISTICA software.[111] The considered tolerance parameter (proportion of variance that is unique to the respective variable) was the default value for minimum acceptable tolerance, which is 0.01. A forward-stepwise search procedure was fixed as the strategy for variable selection. The principle of parsimony (Occam's razor) was taken into account as a strategy for model selection. The quality of the models was determined by examining Wilks' λ parameter (*U* statistic), the square Mahalanobis distance (D^2), the Fisher ratio (*F*), and the corresponding *p* level [*p*(*F*)] as well as the percentage of good classification (accuracy) in the training and test sets. The classification of cases was performed by means of the posterior classification probabilities. By using the models, one compound can then be classified as active, if $\Delta P\% > 0$, being $\Delta P\% = [P(\text{Active}) - P(\text{Inactive})]>100$, or as inactive otherwise. *P*(Active) and *P*(Inactive) are the probabilities with which the equations classify a compound as active or inactive, respectively. Performing the

assessment of the obtained models, the sensibility, the specificity (also known as "hit rate"), the false positive rate (also known as "false alarm rate"), and Matthews' correlation coefficient (*C*), were calculated; and checked in the training and test sets.[112] Finally, leverage approach[93] was used to evaluate the AD of QSAR models. Through of this method it is possible to verify whether a new chemical will lie within the structural model domain. The leverage *h* of a compound measures its influence on the model. That is, leverage used as a *quantitative measure* of the model AD is suitable for evaluating the degree of extrapolation, which represents a sort of compound "distance" from the model experimental space. Leverage values can be calculated for both training compounds and new compounds. In the first case, they are useful for finding training compounds that influence model parameters to a marked extent, resulting in an unstable model. In the second case, they are useful for checking the applicability domain of the model.[94] The warning leverage**,** *h**, is a *critical value* or *cut-off* to consider the prediction made for the model for a specific compounds in dataset. The leverage *h** can be defined 3 *x* p'/n, where *n* is the number of training chemicals and p' is the number of model parameters plus one.[91, 94] Prediction should be considered unreliable for compounds of high leverage value ($h>h^*$). A leverage greater than the warning leverage *h** means that the compound predicted response can be extrapolated from the model, and therefore, the predicted value must be used with great care. Only predicted data for chemicals belonging to the chemical domain of the training set should be proposed. However, this fact can be seeing for two points of view taken into consideration the set of compounds evaluated. For example, when the leverage value of a compound is lower than the critical value, the probability of accordance between predicted and actual values is as high as that for the training set chemicals (good leverage). Conversely, a

high leverage chemical in the test set is structurally distant from the training chemicals (bad leverage), thus it can be considered outside the AD of the model.

Prediction algorithms and ensemble classifier (multi-agent predictor or fusion approach). Here, we used nonstochastic and stochastic linear indices to develop classification-based QSAR models in order to classified molecules as antiprotozoan or inactive compounds. These MDs have a few parameters that it can be "modified" in the calculation process. The number of these *uncertain* parameters depends on what atom-labels (AP scheme) were used for the prediction engine. It would be much more tedious and time-consuming to determine the optimal values for AP [AP:[90] Atomic mass (AP = M), atomic polarizability (AP = P), atomic Mullinken electronegativity (AP = K) and van der Waals atomic volume (AP = V)] *uncertain* parameters. In addition, the number of uncertain parameters also depends on which MDs sets are used to represent the chemical samples. For instance, here every model can be fitted by two kinds of MD sets: 1) non-stochastic MDs (NS), 2) stochastic MDs (SS). To solve the problem, let us use a [2AP+1NS+1SS+1(NS + SS)]-dimensional *fusion approach* (11 models in total), similarity to who early also intruded in protein research.[106]

First, the basic individual classifiers to be generally expressed like $\boldsymbol{C}_{\mathbf{I}}$(NS-AP, SS-AP, NS, SS, NS +SS) and the predicted classification results for a query molecule **M** by each of the individual classifiers can be formulated by,

$$\boldsymbol{C}_{\mathbf{I}}(\text{NS-AP, SS-AP, NS, SS, NS +SS}) \rangle \mathbf{M} = \boldsymbol{C}_{\text{NS-AP, SS-AP, NS, SS, NS +SS}}(\mathbf{M}) \in \mathbf{S} \qquad (\mathbf{12})$$

where the symbol › is an action operator meaning using $\boldsymbol{C}_{\mathbf{I}}$(NS-AP, SS-AP, NS, SS, NS +SS) to classify **M**, **S** representing the union of the two subsets defined (active or inactive). Therefore, the final predicted result should be determined by a fusion approach through the following *voting mechanism*. Now let us introduce an ensemble

classifier $\boldsymbol{C}_{\mathbf{E}}$, which is formed by fusing all set of the basic individual classifiers $\boldsymbol{C}_{\mathbf{I}}$(NS-AP, SS-AP, NS, SS, NS +SS) and can be formulated the follow:

$\boldsymbol{C}_{\mathbf{E}} = \boldsymbol{C}_1(M,NS)\forall \boldsymbol{C}_2(K,NS)\forall \boldsymbol{C}_3(P,NS)\forall \boldsymbol{C}_4(V,NS)\forall \boldsymbol{C}_5$(all AP,NS) …

$\forall \boldsymbol{C}_6$(M,SS)$\forall \boldsymbol{C}_7$(K,SS)$\forall \boldsymbol{C}_8$(P,SS)$\forall \boldsymbol{C}_9$(V,SS)$\forall \boldsymbol{C}_{10}$(all AP,SS) …$\forall \boldsymbol{C}_{11}$(all AP,NS+SS) (**13**)

where the symbol ∀ denotes the fusing operator. Then, the voting *score* for the query molecules **M** belonging to the c^{th} class is given by,

$$\pi_c = \sum_{AP=1}^{4} \sum_{MDs=1}^{2} w_{AP,MDs} \Delta(AP, MDs, S_c) + \sum_{MDs=1}^{3} w_{all-AP,MDs} \Delta(all - AP, MDs, S_c), (c = 1, -1) \quad (\mathbf{14})$$

where $S_c = 1$ is for antiprotozoans and $S_c = -1$ for non-antiprotozoans, $w_{AP,MDs}$ and $w_{all-AP,MDs}$ are the *weight factors* and were set at 1 for simplicity. The delta functions in Eq. **14** is given by,

$$\Delta(AP, MDs, S_c) = \begin{cases} 1 & if\ C_{AP,MDs}(M) \in S_c \\ 0 & otherwise \end{cases} \quad (\mathbf{15})$$

$$\Delta(all - AP, MDs, S_c) = \begin{cases} 1 & if\ C_{P,MDs}(M) \in S_c \\ 0 & otherwise \end{cases} \quad (\mathbf{16})$$

thus the query Molecule **M** is predicted belonging to the class (*c*) or subset **S***c* for which the score of Eq. **14** is the highest; i.e.,

$$\mu = \arg \max_c \{\pi_c\}, \qquad (c = 1, -1) \quad (\mathbf{17})$$

where μ is the argument of *c* that maximize π_c. If there is a tie, then the final predicted result will be randomly assigned (or is take as unclassified) to one of their corresponding subsets although this kind of tie case rarely happens and actually was not observed in the current study.

Chemistry

Instrumental data. Mps were determined in a Stuart Scientific melting point apparatus SMP3. The mps of quinoxalinium salts **5** and **6** as well as those of some of the final products (hydrobromides **9-18**) are not very well defined; these compounds decompose on heating and the observed mps are frequently heating-rate dependent and previous softening is usual. ^{1}H (300 or 400 MHz) and ^{13}C (75 or 100 MHz) NMR spectra were recorded on Varian Unity 300 or Varian Inova 400 spectrometers. The chemical shifts are reported in ppm from TMS (δ scale) but were measured against the solvent signal. *J* values are given in Hz. The assignments have been performed by means of different standard 1D and 2D correlation experiments (NOE, COSY, HMQC and HMBC). Numbering used in the description of NMR spectra of spiro compounds **5** and **6**, and 4-substituted quinoxalinones **7-18** is shown is Scheme; double primed numbers refer to the cyclic secondary amine rings of final compounds **9-18**. Electron impact (EI) and electrospray (ES) mass spectra were obtained at 70 eV on a Hewlett Packard 5973 MSD spectrometer or on a Hewlett Packard 1100 MSD spectrometer, respectively. DC-Alufolien silica gel 60 PF_{254} (Merck, layer thickness 0.2 mm) was used for TLC. Microanalyses were performed by the Departamento de Análisis, Centro de Química Orgánica "Manuel Lora Tamayo", CSIC, Madrid, Spain.

Procedure for the preparation of all chemicals.

2-Bromo-5'-nitro-2'-piperidinoacetanilide (2).- Bromoacetyl bromide (9.08 g, 45 mmol) was dropped (*ca.* 5 min) into a solution of 5-nitro-2-piperidinoaniline (**1**)[70] (8.85 g, 40 mmol) in acetone (150 mL). After 15 min, an additional amount of bromoacetyl bromide (*ca.* 1 mL) was dropped and the mixture stirred for 15 min. The obtained suspension (**2**×HBr) was poured into water (1 L), and the mixture stirred for 30 min. The solid in suspension, collected by filtration, washed with water (4×100 mL) and air-dried was shown to be bromoacetanilide **2** (13.28 g, 97% yield). This compound,

crystallized from ethanol, melts partially and resolidifies at 123-125 °C (decomposition to spiro salt **5**, TLC), showing a further m. p. at 186-190 °C (corresponding to that of salt **5**, see below); ^{1}H NMR ($CDCl_3$): δ 9.40 (s, 1H, NH), 9.20 (d, *J* = 2.7 Hz, 1H, 6'-H), 7.97 (dd, *J* = 8.8, 2.7 Hz, 1H, 4'-H), 7.21 (d, *J* = 8.8 Hz, 1H, 3'-H), 4.09 (s, 2H, 2-H), 2.87 (m, 4H, 2''-, 6''-H), 1.81 (m, 4H, 3''-, 5''-H), 1.62 (m, 2H, 4''-H); ^{13}C NMR ($CDCl_3$): δ 163.46 (C-1), 148.95 (C-2'), 144.16 (C-5'), 132.65 (C-1'), 120.31, 120.01 (C-3', -4'), 114.47 (C-6'), 53.32 (C-2'', -6''), 29.59 (C-2), 26.35 (C-3'', -5''), 23.75 (C-4''); MS (EI): *m/z* (%) 343 (12) ($[M+2]^+$), 341 (12) (M^+), 262 (85), 220 (100), 203 (35), 192 (13), 174 (25), 164 (16), 145 (10), 118 (19). Anal. calcd. for $C_{13}H_{16}BrN_3O_3$ (342.19): C 45.63; H 4.71; N 12.28. Found: C 45.70; H 4.67; N 12.12.

6-Nitro-3-oxo-1,2,3,4-tetrahydroquinoxaline-1-spiro-1'-piperidinium bromide (5).-a) From bromoacetanilide **2**: A solution of anilide **2** (0.68 g, 2.0 mmol) in nitromethane (10 mL) was refluxed for 25 min. After cooling, the insoluble bromide **5** (0.59 g, 87% yield) was collected by filtration, washed with acetone (3x10 mL) and air-dried. b) From tetrahydroquinoxaline-1-spiro-1'-piperidinium chloride **3**: Chloride **3** (prepared[76] by cyclization of 2-chloro analogue of **2**) (7.44 g, 25 mmol) was dissolved in 48% aq. hydrobromic acid and evaporated to dryness. This process was repeated twice and, after addition of acetone (100 mL), the insoluble salt **5** (8.47 g, 99% yield) was collected by filtration, washed with acetone (3 x 40 mL) and air-dried. M. p. 187-192 °C (decomp.) (water); ^{1}H NMR (DMSO-d_6): δ 11.88 (s, 1H, 4-H), 8.30 (d, *J* = 9.0 Hz, 1H, 8-H), 8.13 (dd, *J* = 9.0, 2.6 Hz, 1H, 7-H), 8.01 (d, *J* = 2.6 Hz, 1H, 5-H), 4.89 (s, 2H, 2-H), 4.12 (m, J_{gem} = (-)12.0 Hz, $J_{a,a}$ = 9.6 Hz, 2H, 2'-, 6'-H_a), 3.84 (br d, J_{gem} = (-)12.0 Hz, 2H, 2'-, 6'-H_e), 2.18 (m, 2H) and 1.98-1.50 (m, 4H) (3'-, 4'-, 5'-H); ^{13}C NMR (DMSO-d_6): δ 160.74 (C-3), 148.59 (C-6), 134.98, 133.47 (C-4a, -8a), 122.84 (C-8), 118.38 (C-7), 112.89 (C-5), 61.72 (C-2', -6'), 55.05 (C-2), 19.92 (C-4'), 19.29 (C-3', -

5'); MS (ES+): *m/z* (%) 523 (20) ([2(M-Br)-1]$^+$), 262 (100) ([M-Br]$^+$); MS (EI) of salt **5** is identical to that of bromoalkyl derivative **7** arising from its thermal decomposition. Anal. calcd. for $C_{13}H_{16}BrN_3O_3$ (342.19): C 45.63; H 4.71; N 12.28. Found: C 45.50; H 4.87; N 12.52.

4-Methyl-6-nitro-3-oxo-1,2,3,4-tetrahydroquinoxaline-1-spiro-1'-piperidinium bromide (6).- Spiro chloride **4** (prepared[76] from 2,2'-dichloro-*N*-methyl-5'-nitroacetanilide and piperidine or by cyclization of 2-chloro-*N*-methyl analogue of **2**) (7.79 g, 25 mmol), treated with 48% aq. hydrobromic acid as described for the preparation of salt **5**, afforded the title bromide **6** (7.93 g, 89% yield). M. p. 159-162 °C (decomp.) (ethanol); ^{1}H NMR (DMSO-d_6): δ 8.34 (d, J = 9.0 Hz, 1H, 8-H), 8.23 (dd, J = 9.0, 2.4 Hz, 1H, 7-H), 8.18 (d, J = 2.4 Hz, 1H, 5-H), 4.97 (s, 2H, 2-H), 4.14 (m, J_{gem} = (-)12.1 Hz, $J_{a,a}$ = 9.8 Hz, 2H, 2'-, 6'-H$_a$), 3.90 (br d, J_{gem} = (-)12.1 Hz, 2H, 2'-, 6'-H$_e$), 3.45 (s, 3H, 4-CH$_3$), 2.17 (m, 2H) and 1.93-1.51 (m, 4H) (3'-, 4'-, 5'-H); ^{13}C NMR (DMSO-d_6): δ 160.14 (C-3), 148.93 (C-6), 136.39, 135.34 (C-4a, -8a), 122.62 (C-8), 118.87 (C-7), 112.82 (C-5), 61.48 (C-2', -6'), 55.18 (C-2), 29.69 (4-CH$_3$), 19.93 (C-4'), 19.32 (C-3', -5'); MS (ES+): *m/z* (%) 633 (5) ([2M-Br+2]$^+$), 631 (5) ([2M-Br]$^+$), 276 (100) ([M-Br]$^+$); MS (EI) of salt **6** is identical to that of bromoalkyl derivative **8** arising from its thermal decomposition. Anal. calcd. for $C_{14}H_{18}BrN_3O_3$ (356.22): C 47.20; H 5.09; N 11.80. Found: C 47.48; H 4.87; N 11.62.

4-(5-Bromopentyl)-7-nitro-3,4-dihydro-1*H*-quinoxalin-2-one (7).- A suspension of bromide **5** (6.84 g, 20 mmol) in nitromethane (50 mL) was refluxed for 48 h under argon atmosphere. After cooling, the solid in suspension, collected by filtration, washed with nitromethane (2 x 10 mL) and air-dried, was shown to be the title bromopentyl derivative **7** (6.43 g, 94% yield). Similar results were obtained starting from bromoacetanilide **2**, following the same procedure but without isolation of the

intermediate salt **5**. M. p. 192-195 °C (decomp.) (nitromethane). ^{1}H NMR (DMSO-d_6): δ 10.78 (s, 1H, 1-H), 7.77 (dd, *J* = 9.3, 2.7 Hz, 1H, 6-H), 7.59 (d, *J* = 2.7 Hz, 1H, 8-H), 6.77 (d, *J* = 9.3 Hz, 1H, 5-H), 4.04 (s, 2H, 3-H), 3.53 (t, *J* = 6.7 Hz, 2H, 5'-H), 3.34 (t, *J* = 7.3 Hz, 2H, 1'-H), 1.84 (m, 2H, 4'-H), 1.59 (m, 2H, 2'-H), 1.42 (m, 2H, 3'-H); ^{13}C NMR (DMSO-d_6): δ 163.82 (C-2), 140.20 (C-4a), 136.46 (C-7), 125.57 (C-8a), 120.58 (C-6), 109.66 (C-8), 109.17 (C-5), 51.00 (C-3), 48.96 (C-1'), 35.00 (C-5'), 31.99 (C-24'), 24.92 (C-3'), 23.64 (C-2'); MS (EI): *m/z* (%) 343 (24) ($[M+2]^+$), 341 (24) (M^+), 262 (17), 206 (100), 178 (45), 160 (10), 132 (23), 118 (8). Anal. calcd. for $C_{13}H_{16}BrN_3O_3$ (342.19): C 45.63; H 4.71; N 12.28. Found: C 45.55; H 4.61; N 12.52.

4-(5-Bromopentyl)-1-methyl-7-nitro-3,4-dihydro-1*H*-quinoxalin-2-one (8).- A suspension of bromide **6** (7.12 g, 20 mmol) in nitromethane (50 mL) was refluxed for 24 h under argon. The solvent was then evaporated to dryness and the residue triturated with ethanol (20 mL); the insoluble material was collected by filtration, washed with cold ethanol (2 x 10 mL) and air-dried affording compound **8** (5.91 g, 83% yield). M. p. 118-120 °C (ethanol). ^{1}H NMR (DMSO-d_6): δ 7.88 (dd, *J* = 9.3, 2.4 Hz, 1H, 6-H), 7.69 (d, *J* = 2.4 Hz, 1H, 8-H), 6.85 (d, *J* = 9.3 Hz, 1H, 5-H), 4.12 (s, 2H, 3-H), 3.53 (t, *J* = 6.7 Hz, 2H, 5'-H), 3.37 (t, *J* = 7.6 Hz, 2H, 1'-H), 3.32 (s, 3H, 1-CH_3), 1.84 (m, 2H, 4'-H), 1.58 (m, 2H, 2'-H), 1.43 (m, 2H, 3'-H; ^{13}C NMR (DMSO-d_6): δ 163.27 (C-2), 141.71 (C-4a), 136.85 (C-7), 127.64 (C-8a), 120.79 (C-6), 109.65, 109.51 (C-5, -8), 51.02 (C-3), 49.01 (C-1'), 34.94 (C-5'), 31.92 (C-4'), 28.25 (1-CH_3), 24.87 (C-3'), 23.57 (C-2'); MS (EI): *m/z* (%) 357 (33) ($[M+2]^+$), 355 (33) (M^+), 276 (17), 220 (100), 192 (72), 160 (7), 146 (29), 131 (12), 104 (5)**.** Anal. calcd. for $C_{14}H_{18}BrN_3O_3$ (356.22): C 47.20; H 5.09; N 11.80. Found: C 47.48; H 5.19; N 11.62.

Preparation of 4-[5-(dialkylamino)pentyl]quinoxalin-2-ones hydrobromides 9-18.- For dimethylamino derivatives **9** and **14**, the corresponding bromide (**7** or **8**) (3

mmol) and dimethylamine (7.5 mmol; 1.34 mL of a 5.6 M solution in ethanol) in 1,4-dioxane (100 mL) was heated in an autoclave at 100-110 °C until the starting bromide was consumed (*ca.* 6 h). For cyclic secondary amines derivatives **10**-**13** and **15**-**18**, a mixture of the corresponding bromide (**7** or **8**) (3 mmol) and the required amine (7.5 mmol) in 1,4-dioxane (100 mL) was refluxed until the starting bromide was consumed (5-10 h). After eventual separation (filtration or decantation) of some tars appeared when using dimethylamine or pyrrolidine, dioxane was evaporated to dryness and ethanol (10 mL) and 48% aq. hydrobromic acid (0.5 mL) were added. The mixture was stirred for 2 h and the precipitated hydrobromide collected by filtration, washed with ethanol (2 x 5 mL) and air-dried (83-98% yield).

4-[5-(Dimethylamino)pentyl]-7-nitro-3,4-dihydro-1*H*-quinoxalin-2-one hydrobromide (9).- Yield: 0.98 g (84%); M. p. 204-207 °C (methanol); ^{1}H NMR (DMSO-d_6): δ 10.81 (s, 1H, 1-H), 9.44 (br s, 1H, 5'-NH^+), 7.77 (dd, J = 9.3, 2.7 Hz, 1H, 6-H), 7.60 (d, J = 2.7 Hz, 1H, 8-H), 6.80 (d, J = 9.3 Hz, 1H, 5-H), 4.06 (s, 2H, 3-H), 3.35 (t, J = 7.5 Hz, 2H, 1'-H), 3.03 (m, 2H, 5'-H), 2.75 [s, 6H, $N(CH_3)_2$], 1.61 (m, 4H, 2'-, 4'-H), 1.33 (m, 2H, 3'-H); ^{13}C NMR (DMSO-d_6): δ 163.86 (C-2), 140.27 (C-4a), 136.43 (C-7), 125.59 (C-8a), 120.60 (C-6), 109.66, 109.30 (C-5, -8), 56.34 (C-5'), 51.08 (C-3), 48.84 (C-1'), 42.09 [$N(CH_3)_2$], 24.03, 23.44, 23.09 (C-2', -3', -4'). MS (ES+): *m/z* (%) 695 (12) ($[2M-Br+2]^+$), 693 (12) ($[2M-Br]^+$), 308 (20) ($[M-Br+1]^+$), 307 (100) ($[M-Br]^+$). Anal. calcd. for $C_{15}H_{23}BrN_4O_3$ (387.27): C 46.52; H 5.99; N 14.47. Found: C 46.50; H 5.77; N 14.21.

7-Nitro-4-(5-pyrrolidinopentyl)-3,4-dihydro-1*H*-quinoxalin-2-one hydrobromide (10).- Yield: 1.22 g (98%); M. p. 233-235 °C (decomp.) (water); ^{1}H NMR (DMSO-d_6): δ 10.81 (s, 1H, 1-H), 9.60 (br s, 1H, 1''-H), 7.77 (dd, J = 9.3, 2.7 Hz, 1H, 6-H), 7.60 (d, J = 2.7 Hz, 1H, 8-H), 6.80 (d, J = 9.3 Hz, 1H, 5-H), 4.06 (s, 2H, 3-H),

3.49 (br s, 2H, 2''-, 5''-H_A), 3.35 (t, *J* = 7.4 Hz, 2H, 1'-H), 3.10 (m, 2H, 5'-H), 2.97 (br s, 2H, 2''-, 5''-H_B), 1.95 (br s, 2H) and 1.87 (br s, 2H) (3''-, 4''-H), 1.63 (m, 4H, 2'-, 4'-H), 1.33 (m, 2H, 3'-H). ^{13}C NMR (DMSO-d_6): δ 163.79 (C-2), 140.22 (C-4a), 136.38 (C-7), 125.54 (C-8a), 120.59 (C-6), 109.62, 109.29 (C-5, -8), 53.58 (C-5'), 52.94 (C-2'', -5''), 51.05 (C-3), 48.85 (C-1'), 24.85, 23.96, 23.23 (C-2', -3', -4'), 22.61 (C-3'', -4''). MS (ES+): *m/z* (%) 747 (14) ([2M-Br+2]$^+$), 745 (13) ([2M-Br]$^+$), 334 (23) ([M-Br+1]$^+$), 333 (100) ([M-Br]$^+$). Anal. calcd. for $C_{17}H_{25}BrN_4O_3$ (413.31): C 49.40; H 6.10; N 13.56. Found: C 49.50; H 6.37; N 13.72.

7-Nitro-4-(5-piperidinopentyl)-3,4-dihydro-1*H*-quinoxalin-2-one hydrobromide (11).- Yield: 1.24 g (97%); M. p. 246-248 °C (decomp.) (methanol); ^{1}H NMR (DMSO-d_6): δ 10.81 (s, 1H, 1-H), 9.08 (br s, 1H, 1''-H), 7.78 (dd, *J* = 9.0, 2.7 Hz, 1H, 6-H), 7.60 (d, *J* = 2.7 Hz, 1H, 8-H), 6.80 (d, *J* = 9.0 Hz, 1H, 5-H), 4.06 (s, 2H, 3-H), 3.37 (m, 4H, 1'-H, 2''-, 6''-H_e), 3.00 (m, 2H, 5'-H), 2.83 (m, 2H, 2''-, 6''-H_a), 1.67 (m, 9H, 2'-, 4'-, 3''-, 5''-H, 4''-H_A), 1.33 (m, 3H, 3'-H, 4''-H_B); ^{13}C NMR (DMSO-d_6): δ 163.85 (C-2), 140.26 (C-4a), 136.43 (C-7), 125.59 (C-8a), 120.59 (C-6), 109.66, 109.28 (C-5, -8), 55.59 (C-5'), 51.95 (C-2'', -6''), 51.06 (C-3), 48.81 (C-1'), 24.02, 23.31, 22.93 (C-2', -3', -4'), 22.45 (C-3'', -5''), 21.34 (C-4''). MS (ES+): *m/z* (%) 775 (8) ([2M-Br+2]$^+$), 773 (8) ([2M-Br]$^+$), 348 (25) ([M-Br+1]$^+$), 347 (100) ([M-Br]$^+$). Anal. calcd. for $C_{18}H_{27}BrN_4O_3$ (427.34): C 50.59; H 6.37; N 13.11. Found: C 50.50; H 6.47; N 13.32.

4-(5-Azepanylpentyl)-7-nitro--3,4-dihydro-1*H*-quinoxalin-2-one hydrobromide (12).- Yield: 1.28 g (97%); M. p. 235-237 °C (decomp.) (methanol); ^{1}H NMR (DMSO-d_6): δ 10.82 (s, 1H, 1-H), 9.13 (br s, 1H, 1''-H), 7.79 (dd, *J* = 9.0, 2.7 Hz, 1H, 6-H), 7.61 (d, *J* = 2.7 Hz, 1H, 8-H), 6.80 (d, *J* = 9.0 Hz, 1H, 5-H), 4.06 (s, 2H, 3-H), 3.35 (m, 4H, 1'-H, 2''-, 7''-H_A), 3.06 (m, 4H, 5'-H, 2''-, 7''-H_B), 1.90-1.50 (m, 12H,

2’-, 4’-, 3’’-, 4’’-, 5’’-, 6’’-H), 1.33 (m, 2H, 3’-H); ^{13}C NMR (DMSO-d_6): δ 163.83 (C-2), 140.26 (C-4a), 136.42 (C-7), 125.58 (C-8a), 120.58 (C-6), 109.64, 109.27 (C-5, -8), 56.06 (C-5’), 53.53 (C-2’’, -7’’), 51.06 (C-3), 48.82 (C-1’), 25.94 (C-3’’, -6’’), 24.04, 23.32, 23.29 (C-2’, -3’, -4’), 22.86 (C-4’’, -5’’). MS (ES+): *m/z* (%) 803 (18) ([2M-Br+2]$^+$), 801 (17) ([2M-Br]$^+$), 362 (24) ([M-Br+1]$^+$), 361 (100) ([M-Br]$^+$). Anal. calcd. for $C_{19}H_{29}BrN_4O_3$ (441.36): C 51.70; H 6.62; N 12.69. Found: **C 51.98; H 6.67; N 12**.62.

7-Nitro-4-[5-(1,2,3,4-tetrahydroisoquinolin-2-yl)pentyl]-3,4-dihydro-1*H*-quinoxalin-2-one hydrobromide (13).- Yield: 1.34 g (94%); M. p. 196-198 °C (decomp.) (0.5 M aq HBr); ^{1}H NMR (DMSO-d_6): δ 10.83 (s, 1H, 1-H), 9.73 (br s, 1H, 2’’-H), 7.79 (dd, *J* = 9.3, 2.7 Hz, 1H, 6-H), 7.61 (d, *J* = 2.7 Hz, 1H, 8-H), 7.32-7.16 (m, 4H, 5’’-, 6’’-, 7’’-, 8’’-H), 6.82 (d, *J* = 9.3 Hz, 1H, 5-H), 4.55 [br d, *J* = (-)15.3 Hz, 1’’-H_A], 4.29 [br dd, *J* = (-)15.3, 8.3 Hz, 1’’-H_B], 4.07 (s, 2H, 3-H), 3.70 (m, 1H, 3’’-H_A), 3.35 (t, *J* = 7.2 Hz, 2H, 1’-H), 3.30-2.95 (m, 5H, 5’-,4’’-H, 3’’-H_B), 1.79 (m, 2H, 4’-H), 1.63 (m, 2H, 2’-H), 1.38 (m, 2H, 3’-H). ^{13}C NMR (DMSO-d_6): δ 163.73 (C-2), 140.19 (C-4a), 136.44 (C-7), 131.25, 128.48, 126.59, 126.55 (C-5’’, -6’’, -7’’, -8’’), 128.31, 127.64 (C-4’’a, -8’’a), 125.55 (C-8a), 120.48 (C-6), 109.62 (C-8), 109.25 (C-5), 54.90 (C-5’), 51.71 (C-1’’), 51.05 (C-3), 48.81 (C-1’), 48.76 (C-3’’), 24.78 (C-4’’), 24.00 (C-2’), 23.21 (C-3’), 23.03 (C-4’). MS (ES+): *m/z* (%) 871 (6) ([2M-Br+2]$^+$), 869 (6) ([2M-Br]$^+$), 396 (28) ([M-Br+1]$^+$), 395 (100) ([M-Br]$^+$). Anal. calcd. for $C_{22}H_{27}BrN_4O_3$ (475.38): C 55.58; H 5.72; N 11.79. Found: C 55.50; H 5.67; N 11.52.

4-[5-(Dimethylamino)pentyl]-1-methyl-7-nitro-3,4-dihydro-1*H*-quinoxalin-2-one hydrobromide (14).- Yield: 1.00 g (83%); M. p. 182-184 °C (decomp.) (ethanol); ^{1}H NMR (DMSO-d_6): δ 9.38 (br s, 1H, 5’-NH$^+$), 7.88 (dd, *J* = 9.3, 2.4 Hz, 1H, 6-H), 7.70 (d, *J* = 2.4 Hz, 1H, 8-H), 6.88 (d, *J* = 9.3 Hz, 1H, 5-H), 4.13 (s, 2H, 3-H), 3.38 (t, *J*

= 7.1 Hz, 2H, 1'-H), 3.32 (s, 3H, 1-CH_3), 3.03 (m, 2H, 5'-H), 2.74 [s, 6H, $N(CH_3)_2$], 1.63 (m, 4H, 2'-, 4'-H), 1.34 (m, 2H, 3'-H); ^{13}C NMR (DMSO-d_6): δ 163.35 (C-2), 141.81 (C-4a), 136.88 (C-7), 127.70 (C-8a), 120.85 (C-6), 109.78, 109.61 (C-5, -8), 56.35 (C-5'), 51.13 (C-3), 48.91 (C-1'), 42.10 [$N(CH_3)_2$], 28.30 (1-CH_3), 23.99, 23.42, 23.09 (C-2', -3', -4'); MS (ES+): *m/z* (%) 723 (8) ([2M-Br+2]$^+$), 721 (8) ([2M-Br]$^+$), 322 (20) ([M-Br+1]$^+$), 321 (100) ([M-Br]$^+$). Anal. calcd. for $C_{16}H_{25}BrN_4O_3$ (401.30): C 47.89; H 6.28; N 13.96. Found: C 47.64; H 6.47; N 13.92.

1-Methyl-7-nitro-4-(5-pyrrolidinopentyl)-3,4-dihydro-1*H*-quinoxalin-2-one hydrobromide (15).- Yield: 1.13 g (88%); M. p. 206-208 °C (decomp.) (ethanol).; ^{1}H NMR (DMSO-d_6): δ 9.46 (br s, 1H, 1''-H), 7.90 (dd, *J* = 9.0, 2.4 Hz, 1H, 6-H), 7.72 (d, *J* = 2.4 Hz, 1H, 8-H), 6.87 (d, *J* = 9.0 Hz, 1H, 5-H), 4.14 (s, 2H, 3-H), 3.50 (br s, 2H, 2''-, 5''-H_A), 3.38 (t, *J* = 7.3 Hz, 2H, 1'-H), 3.33 (s, 3H, 1-CH_3), 3.10 (m, 2H, 5'-H), 2.96 (br s, 2H, 2''-, 5''-H_B), 1.97 (br s, 2H) and 1.83 (br s, 2H) (3''-, 4''-H), 1.61 (m, 4H, 2'-, 4'-H), 1.35 (m, 2H, 3'-H).); ^{13}C NMR (DMSO-d_6): δ 163.34 (C-2), 141.79 (C-4a), 136.84 (C-7), 127.67 (C-8a), 120.87 (C-6), 109.76, 109.63 (C-5, -8), 53.62 (C-5'), 53.01 (C-2'', -5''), 51.12 (C-3), 48.94 (C-1'), 28.30 (1-CH_3), 24.88, 23.96, 23.24 (C-2', -3', -4'), 22.58 (C-3'', -4''); MS (ES+): *m/z* (%) 775 (15) ([2M-Br+2]$^+$), 773 (15) ([2M-Br]$^+$), 348 (24) ([M-Br+1]$^+$), 347 (100) ([M-Br]$^+$). Anal. calcd. for $C_{18}H_{27}BrN_4O_3$ (427.34): C 50.59; H 6.37; N 13.11. Found: C 50.33; H 6.61; N 13.33.

1-Methyl-7-nitro-4-(5-piperidinopentyl)-3,4-dihydro-1*H*-quinoxalin-2-one hydrobromide (16).- Yield: 1.24 g (94%); M. p. 214-216 °C (decomp.) (ethanol); ^{1}H NMR (DMSO-d_6): δ 9.24 (br s, 1H, 1''-H), 7.88 (dd, *J* = 9.3, 2.4 Hz, 1H, 6-H), 7.69 (d, *J* = 2.4 Hz, 1H, 8-H), 6.88 (d, *J* = 9.3 Hz, 1H, 5-H), 4.13 (s, 2H, 3-H), 3.37 (m, 4H, 1'-H, 2''-, 6''-H_e), 3.31 (s, 3H, 1-CH_3), 3.00 (m, 2H, 5'-H), 2.85 (m, 2H, 2''-, 6''-H_a), 1.90-1.50 (m, 9H, 2'-, 4'-, 3''-, 5''-H, 4''-H_A), 1.33 (m, 3H, 3'-H, 4''-H_B); ^{13}C NMR

(DMSO-d_6): δ 163.34 (C-2), 141.79 (C-4a), 136.84 (C-7), 127.68 (C-8a), 120.87 (C-6), 109.76, 109.62 (C-5, -8), 55.58 (C-5'), 51.93 (C-2'', -6''), 51.12 (C-3), 48.91 (C-1'), 28.30 (1-CH_3), 23.96, 23.31, 22.89 (C-2', -3', -4'), 22.43 (C-3'', -5''), 21.34 (C-4''); MS (ES+): *m/z* (%) 803 (15) ($[2M-Br+2]^+$), 801 (13) ($[2M-Br]^+$), 362 (24) ($[M-Br+1]^+$), 361 (100) ($[M-Br]^+$). Anal. calcd. for $C_{19}H_{29}BrN_4O_3$ (441.36): C 51.70; H 6.62; N 12.69. Found: C 51.57; H 6.67; N 12.45.

4-(5-Azepanylpentyl)-1-methyl-7-nitro--3,4-dihydro-1*H*-quinoxalin-2-one hydrobromide (17).- Yield: 1.24 g (91%); M. p. 223-225 °C (decomp.) (ethanol); ^{1}H NMR (DMSO-d_6): δ 9.11 (br s, 1H, 1''-H), 7.90 (dd, *J* = 9.0, 2.6 Hz, 1H, 6-H), 7.72 (d, *J* = 2.6 Hz, 1H, 8-H), 6.87 (d, *J* = 9.0 Hz, 1H, 5-H), 4.13 (s, 2H, 3-H), 3.39 (m, 4H, 1'-H, 2''-, 7''-H_A), 3.33 (s, 3H, 1-CH_3), 3.06 (m, 4H, 5'-H, 2''-, 7''-H_B), 1.90-1.50 (m, 12H, 2'-, 4'-, 3''-, 4''-, 5''-, 6''-H), 1.33 (m, 2H, 3'-H); ^{13}C NMR (DMSO-d_6): δ 163.35 (C-2), 140.81 (C-4a), 136.87 (C-7), 127.70 (C-8a), 120.87 (C-6), 109.77, 109.62 (C-5, -8), 56.07 (C-5'), 53.57 (C-2'', -7''), 51.13 (C-3), 48.92 (C-1'), 28.30 (1-CH_3), 25.93 (C-3'', -6''), 24.01, 23.31 (2C) (C-2', -3', -4'), 22.89 (C-4'', -5''). MS (ES+): *m/z* (%) 831 (13) ($[2M-Br+2]^+$), 829 (12) ($[2M-Br]^+$), 376 (28) ($[M-Br+1]^+$), 375 (100) ($[M-Br]^+$). Anal. calcd. for $C_{20}H_{31}BrN_4O_3$ (455.39): C 52.75; H 6.86; N 12.30. Found: C 52.49; H 6.59; N 12.51.

1-Methyl-7-nitro-4-[5-(1,2,3,4-tetrahydroisoquinolin-2-yl)pentyl]-3,4-dihydro-1*H*-quinoxalin-2-one hydrobromide (18).- Yield: 1.29 g (88%); M. p. 184-187 °C (decomp.) (methanol); ^{1}H NMR (DMSO-d_6): δ 9.95 (br s, 1H, 2''-H), 7.89 (dd, *J* = 9.1, 2.4 Hz, 1H, 6-H), 7.70 (d, *J* = 2.4 Hz, 1H, 8-H), 7.32-7.14 (m, 4H, 5''-, 6''-, 7''-, 8''-H), 6.90 (d, *J* = 9.1 Hz, 1H, 5-H), 4.54 (br s, 1''-H_A), 4.34 (br s, 1''-H_B), 4.15 (s, 2H, 3-H), 3.71 (m, 1H, 3''-H_A), 3.41 (t, *J* = 7.1 Hz, 2H, 1'-H), 3.32 (s, 3H, 1-CH_3), 3.30-2.95 (m, 5H, 5'-,4''-H, 3''-H_B), 1.83 (m, 2H, 4'-H), 1.63 (m, 2H, 2'-H), 1.40 (m, 2H, 3'-H).

^{13}C NMR (DMSO-d_6): δ 163.35 (C-2), 141.80 (C-4a), 136.86 (C-7), 131.31 (CH), 128.58 (CH), 128.41 (C_{ipso}), 127.74 (C_{ipso}), 127.68 (C_{ipso}), 126.68 (CH), 126.63 (CH) (C-8a, -4''a, -5'', -6'', -7'', -8'', -8''a), 120.88 (C-6), 109.76 (C-8), 109.64 (C-5), 54.95 (C-5'), 51.81 (C-1''), 51.16 (C-3), 48.93 (C-1'), 48.85 (C-3''), 28.31 (1-CH_3), 24.87 (C-4''), 24.00 (C-2'), 23.26 (C-3'), 23.13 (C-4'). MS (ES+): *m/z* (%) 897 (7) ([2M-Br+2]$^+$), 895 (7) ([2M-Br]$^+$), 410 (29) ([M-Br+1]$^+$), 409 (100) ([M-Br]$^+$). Anal. calcd. for $C_{23}H_{29}BrN_4O_3$ (489.41): C 56.45; H 5.97; N 11.45. Found: C 56.57; H 6.21; N 11.69.

Wet *Evaluation: Pharmacological Assays.*

Determination of *in vitro* trichomonacidal activity. The biological activity was assayed on *Trichomonas vaginalis* JH31A #4 Ref. No. 30326 (ATCC, MD, USA) in modified Diamond medium supplemented with equine serum and grown at 37 ˚C (5% CO_2). The compounds were added to the cultures at several concentrations (100, 10, and 1 μg/mL) after 6 h of the seeding (0 h). Viable protozoa were assessed at 24 and 48 h after incubation at 37 ˚C by using the Neubauer chamber. Metronidazole (Sigma-Aldrich SA, Spain) was used as reference drug at concentrations of 2, 1, 0.5 μg/mL. Cytocidal and cytostatic activities were determined by calculation of percentages of cytocidal (%C) and cytostatic activities (%CA), in relation to controls as previously reported.[113, 114]

***T. cruzi* epimastigote susceptibility assay**. For this *in vitro* test,[79, 115] CL strain parasites (clone CL-B5) stably transfected with the *Escherichia coli* β-galactosidase gene (*LacZ*) were used. The epimastigotes were grown at 28º C in liver infusion tryptose broth (LIT) with 10% foetal bovine serum (FBS), penicillin and streptomycin and harvested during the exponential growth phase. The screening assay was performed in 96-well microplates (Sarstedt, Sarstedt, Inc.) with cultures that had not reached the stationary phase. Briefly, epimastigotes form, CL strain, was seeded at concentration of

1×105 per milliliter in 200 µl media. The plates were then incubated at 28º C for 72 hours with various concentrations of the drugs (100, 10 and 1 µg/mL), at which time 50 µl of CPRG solution was added to give a final concentration of 200 µM. The plates were incubated at 37º C for 6 hrs and absorbances were then read at 595 nm. Each concentration was tested in triplicate and in order to avoid drawback, medium, negative and drug controls were used in each test. The anti-epimastigote percentage (%AE) was calculated as follows: %AE = [(AE-AEB)/(AC-ACB)] x 100, where AE = absorbance of experimental group; AEB = blank of compounds; AC = Absorbance of control group; ACB = blank of culture medium. Stock solutions of the compounds to be assayed were prepared in DMSO, with the final concentration in a water/DMSO mixture never exceeding 0.2% of the latter solvent.[79, 115] Nifurtimox was used as reference drug.

***In vitro* cytotoxicity on macrophage cells.** Murine J774 macrophages were grown in plastic 25 µl flasks in (RPMI)-1640 medium (Sigma) supplemented with 20% heat inactivated (30 min, 56^{o}C) foetal calf serum (FCS) and 100 IU penicillin/mL + 100 µg/mL streptomycin, in a humidified 5% CO2/95% air atmosphere at 37 ºC and subpassaged once a week. J774 macrophages were seeded (70,000 cells/well) in 96-well flat-bottom microplates (Nunc) with 200 µl of medium. The cells were allowed to attach for 24 h at 37ºC and then exposed to the compounds (dissolved in DMSO, maximal final concentration of solvent was 0.2%) for another 24 h. Afterwards, the cells were washed with PBS and incubated (37ºC) with 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT) 0.4 mg/mL for 60 min. MTT solution was removed and the cells solubilized in DMSO (100 µl). The extent of reduction of MTT to formazan within cells was quantified by measurement of OD595.[116] Each concentration was assayed three times and six cell growth controls were used in each test. The assays were performed in duplicate. Nifurtimox cytotoxicty was also determined. Cytotoxic

percentages (%C) were determined as follows: %C = [1-(ODp-ODpm)/(ODc-ODm)]×100, where ODp represents the mean OD595 value recorded for wells with macrophages containing different doses of product; ODpm represents the mean OD595 value recorded for different concentrations of product in medium; ODc represents the mean OD595 value recorded for wells with macrophages and no product (growth controls), and ODm represents the mean OD595 value recorded for medium/control wells. The 50% cytotoxic dose (CD_{50}) was defined as the concentration of drug that decreases OD595 up to 50% of that in control cultures.[79]

Efficacy studies with *Toxoplasma gondii* tachyzoites. The efficacy of chemicals were tested against tachyzoites form of *Toxoplasma gondii*.[102, 103] Tachyzoites ($1x10^6$) were settled in ependorf microtubes (500 µl, Axygen Scientific), and exposed to compounds **9-18** for four hours at room temperature in order to evaluate the viability of the parasites. One hundred and fifty tachyzoites were counted and the viability percentage was taken with trypan blue exclusion method by counting the number of living tachyzoites

All chemicals were first dissolved in dimethyl sulfoxide [DMSO, sigma, 99.5% (GC)], and then diluted in BME (basal medium eagle) Sigma-Aldrich. The compounds were assayed in the range of 1 mM, 500 µM, 200 µM, 100 µM. The final concentration of DMSO did not exceed 0.2% which caused no damage to the parasite. Later, Balb c mice were used for parasite infections maintained in an animal facility with regulated environment conditions of temperature, humidity and filtered air. Management was performed according to the country official norm NOM-062-ZOO-1999 for the production, care and use of laboratory animals (Mexico). Finally, maintenance and purification of *Toxoplasma gondii* tachyzoites -RH strain tachyzoites- were maintained by i.p. passages in female Balb/c mice. After cervical dislocation, parasites were

recovered by i.p. exudates after a peritoneal washing with PBS 8138 mM NaCl, 1.1 mM K_2PO_4, 0.1 mM Na_2HPO_4 and 2.7 mM KCl, pH 7.2) and purified by filtration through 5μm pore polycarbonate membranes (Millipore Co, Bedford, MN).[102, 103]

Ferriprotoporphyrin (FP) IX biocrystallization inhibition test (FBIT). The procedure for testing FP biocrystallization was performed according to the method of Deharo *et al*.[104] In a normal non-sterile flat bottom 96-well plate at 37 °C for 18–24 h it was placed a mixture containing either 50 μl of drug solution (from 5 to 0.0125 mg/mL) or 50 μl of solvent (for control), 50 μl of 0.5 mg/mL of haemin chloride (Sigma H 5533) freshly dissolved in DMSO and 100 μl of 0.5 M sodium acetate buffer pH 4.4. The final pH of the mixture was in the range 5–5.2. The following order of addition was followed: first the haemin chloride solution, second the buffer, and finally the solvent or the solution of drug. The plate was then centrifuged at 1600 × g for 5 min. The supernatant was discarded by vigorously flipping of the plate upside down the plate twice. The remaining pellet was resuspended with 200 μl of DMSO to remove unreacted FP. The plate was then centrifuged once again and the supernatant similarly discarded. The pellet, consisting of precipitate of β-hematin, was dissolved in 150 μl of 0.1M NaOH for direct (in the same plate) spectroscopic quantification at 405 nm with a micro-ELISA (Enzyme-Linked Immunosorbent Assay) reader (Titertek Multiskan MCC/340). The percentage of inhibition of FP biocrystallization was calculated as follows: Inhibition (%) = 100 × [(O.D. control − O.D. drug)/ O.D. control], where O.D. represents the mean of optical density for either controls or drugs.[65, 104] IC_{50} values were determined using the *TREND* function of Software Excel.

Assessment of antimalarial activity *in vitro* by a semiautomated microdilution technique. A rapid, semiautomated microdilution method was developed for measuring the activity of potential antimalarial drugs against cultured intraerythrocytic asexual

forms of the human malaria parasite *Plasmodium falciparum*.[105] Microtitration plates were used to prepare serial dilutions of the compounds to be tested. Parasites (strain 3D7), obtained from continuous stock cultures, were subcultured in these plates for 42 h. Inhibition of uptake of a radiolabeled nucleic acid precursor by the parasites served as the indicator of antimalarial activity.[105] Chloroquine was used as antimalarial reference drug in this assay.

Supporting Information Available: The complete list of compounds used in training and test sets, as well as their a posteriori classification according to obtained models is available free of charge via the Internet at http://pubs.acs.org.

Acknowledgements: One of the authors (M-P. Y) thanks the program '*Estades Temporals per a Investigadors Convidats*' for a fellowship to work at Valencia University (2008). CAMD-BIR Unit gives thanks to the research project called: "*Strengthening Postgraduate Education and Research in Pharmaceutical Sciences*". This project is funded by the Flemish Interuniversity Council (VLIR) of Belgium. NR was supported by fellowship from CONACYT (Apoyos Integrales para la Formación de Doctores en Ciencias) and DGAPA (PROFIP) UNAM (México). We are grateful to CONACYT (México) for the grant No. 60864 (to RM) which partly supported the antitoxoplasma study. Finally, but not least, the authors acknowledge also the partial financial support from Spanish *"Comisión Interministerial de Ciencia y Tecnología"* (CICYT) (Project Reference: SAF2006-04698).

References and Notes

1. Renslo, A. R.; McKerrow, J. H. Drug Discovery and Development for Neglected Parasitic Diseases. *Nature Chem. Biol.* **2006,** 2, 701-710.
2. Ridley, R. G. Medical Need, Scientific Opportunity and the Drive for Antimalarial Drugs. *Nature* **2002,** 415, 686-693.
3. Baneth, G.; Shawb, S. E. Chemotherapy of Canine Leishmaniosis. *Vet. Parasitol.* **2002,** 106, 315–324.
4. Dardonville, C.; Brun, R. Bisguanidine, Bis(2-aminoimidazoline), and Polyamine Derivatives as Potent and Selective Chemotherapeutic Agents against Trypanosoma brucei rhodesiense. Synthesis and in Vitro Evaluation. *J. Med. Chem.* **2004,** 47, 2296-2307.
5. Roldos, V.; Nakayama, H.; Rolón, M.; Montero-Torres, A.; Trucco, F.; Torres, S.; Vega, M. V.; Marrero-Ponce, Y.; Heguaburu, V.; Yaluff, G.; Gómez-Barrio, A.; Sanabria, L.; Ferreira, M. E.; de Arias, M. A.; Pandolfi, E. Activity of a Hydroxybibenzyl Bryophyte Constituent Against Leishmania spp. and Trypanosoma cruzi: In silico, in vitro and in vivo Activity Studies. *Eur. J. Med. Chem.* **2007,** doi:10.1016/j.ejmech.2007.11.007.
6. Gardner, T. B.; Hill, D. R. Treatment of Giardiasis. *Clinic. Microbiol. Rev.* **2001,** 14, 114–128.
7. Jarroll, E. L.; Sener, K. Potential Drug Targets in Cyst-Wall Biosynthesis by Intestinal Protozoa. *Drug Res. Updates* **2003,** 6, 239-246.
8. WHO's TDR Website. http://www.who.int/tdr/
9. Chavalitshewinkoon-Petmitr, P.; Ramdja, M.; Kajorndechakiat, S.; Ralph, R. K.; Denny, W. A.; Wilairat, P. In vitro Susceptibility of Trichomonas vaginalis to AT-Specific Minor Groove Binding Drugs. *J. Antimicrobial Chem.* **2003,** 52, 287-289.
10. Zuther, E.; Johnson, J. J.; Haselkorn, R.; McLeod, R.; Gornicki, P. Growth of Toxoplasma gondii is Inhibited by Aryloxyphenoxypropionate Herbicides Targeting Acetyl-CoA Carboxylase. *Proc. Natl. Acad. Sci. USA* **1999,** 96, 13387–13392.
11. McKerrow, J. H. Designing Drugs for Parasitic Diseases of the Developing world. *PLoS Med.* **2005,** 2, e210.
12. Mackey, Z. B.; Baca, A. M.; Mallari, J. P.; Apsel, B.; Shelat, A.; Hansell, E. J.; Chiang, P. K.; Wolff, B. W.; Guy, K. R.; Williams, J.; McKerrow, J. H. Discovery of Trypanocidal Compounds by Whole Cell HTS of Trypanosoma brucei. *Chem. Biol. Drug Des.* **2006,** 67, 355-363.
13. St. George, S.; Bishop, J. V.; Titus, R. G.; Selitrennikoff, C. P. Novel Compounds Active Against Leishmania major. *Antimicrob. Agents Chemother.* **2006,** 50, 474–479.
14. Weisman, J. L.; Liou, A. P.; Shelat, A. A.; Cohen, F. E.; Guy, R. K.; DeRisi, J. L. Searching for New Antimalarial Therapeutics Amongst Known Drugs. *Chem. Biol. Drug Des.* **2006,** 67, 409-416.
15. Chong, C. R.; Chen, X.; Shi, L.; Liu, J. O.; Sullivan, D. J. A clinical Drug Library Screen Identifies Astemizole as an Antimalarial Agent. *Nat. Chem. Biol.* **2006,** 2, 415-416.
16. Watson, C. Predictive *in silico* Models in Drug Discovery. *Biosilico* **2003,** 1, 83-84.
17. Xu, J.; Hagler, A. Chemoinformatics and Drug Discovery. *Molecules* **2002,** 7, 566-700.

18. Estrada, E.; Peña, A.; Garcia-Domenech, R. Designing Sedative/Hypnotic Compounds From a Novel Substructural Graph-Theoretical Approach. *J. Comput. Aided Mol. Des.* **1998,** 12, 583-595.
19. Estrada, E.; Uriarte, E.; Montero, A.; Teijeira, M.; Santana, L.; De Clercq, E. A Novel Approach for the Virtual Screening and Rational Design of Anticancer Compounds. *J. Med. Chem.* **2000,** 43, 1975-1985.
20. Davies, J. W.; Glick, M.; Jenkins, J. L. Streamlining Lead Discovery by Aligning in silico and High-Throughput Screening. *Curr. Opi. Stru. Biol.* **2006,** 10, 343-351.
21. Ehrman, T. M.; Barlow, D. J.; Hylands, P. J. Phytochemical Databases of Chinese Herbal Constituents and Bioactive Plant Compounds with Known Target Specificities. *J. Chem. Inf. Model.* **2007,** 47, 254-263.
22. EC2003.2003/15/EC. Commission Directive of 27 February 2003 amending Council Directive 76/768/EEC on the Approximation of Laws of the Member States Relating to Cosmetic Products In Off. J. Eur. Union, Vol. L66, pp 26-35.
23. http://www.europa.eu.int/comm/environment/chemicals/index.html.
24. Marrero-Ponce, Y.; Romero, V. *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* version 1.0; Central University of Las Villas: Santa Clara, Villa Clara, 2002.
25. *STATISTICA (data analysis software system), StatSoft, Inc. version 6. www.statsoft.com.*, 2001.
26. van de Waterbeemd, H. Discriminant Analysis for Activity Prediction. In *Chemometric Methods in Molecular Design*, van de Waterbeemd, H., Ed. VCH Publishers: Weinheim, 1995; pp 265-288.
27. Gonzalez-Diaz, H.; Uriarte, E.; Ramos de Armas, R. Predicting Stability of Arc Repressor Mutants with Protein Stochastic Moments. *Bioorg. Med. Chem.* **2005,** 13, 323-331.
28. Garcia-Garcia, A.; Galvez, J.; de Julian-Ortiz, J. V.; Garcia-Domenech, R.; Munoz, C.; Guna, R.; Borras, R. New Agents Active Against Mycobacterium avium Complex Selected by Molecular Topology: a Virtual Screening Method. *J. Antimicrob. Chemother.* **2004,** 53, 65-73.
29. Cronin, M. T.; Aptula, A. O.; Dearden, J. C.; Duffy, J. C.; Netzeva, T. I.; Patel, H.; Rowe, P. H.; Schultz, T. W.; Worth, A. P.; Voutzoulidis, K.; Schuurmann, G. Structure-Based Classification of Antibacterial Activity. *J. Chem. Inf. Comput. Sci.* **2002,** 42, 869-878.
30. Marrero-Ponce, Y. Total and Local Quadratic Indices of the Molecular Pseudograph´s Atom Adjacency Matrix: Applications to the Prediction of Physical Properties of Organic Compounds. *Molecules* **2003,** 8, 687-726.
31. Marrero-Ponce, Y. Total and Local (Atom and Atom-Type) Molecular Quadratic Indices: Significance-Interpretation, Comparison to Other Molecular Descriptors, and QSPR/QSAR Applications. *Bioorg. Med. Chem.* **2004,** 12, 6351-6369.
32. Marrero Ponce, Y. 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.* **2004,** 44, 2010-26.
33. Marrero-Ponce, Y. Linear indices of the "molecular pseudograph's atom adjacency matrix": definition, significance-interpretation, and application to QSAR analysis

of flavone derivatives as HIV-1 integrase inhibitors. *J. Chem. Inf. Comput. Sci.* **2004,** 44, 2010-26.
34. Marrero-Ponce, Y.; Huesca-Guillen, A.; Ibarra-Velarde, F. 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)* **2005,** 717, 67-79.
35. Marrero-Ponce, Y.; Torrens, F.; García-Domenech, R.; Ortega-Broche, S. E.; Romero Zaldivar, V. Novel 2D TOMOCOMD-CARDD Descriptors: Atom-based Stochastic and non-Stochastic Bilinear Indices and their QSPR Applications. *J. Math. Chem.* **2006**, DOI 10.1007/s10910-008-9389-0.
36. 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. Predicting Antitrichomonal Activity: A Computational Screening Using Atom-Based Bilinear Indices and Experimental Proofs. *Bioorg. Med. Chem.* **2006,** 14, 6502–6524.
37. Marrero Ponce, Y.; Cabrera Perez, M. A.; Romero Zaldivar, V.; Gonzalez Diaz, H.; Torrens, F. A new topological descriptors based model for predicting intestinal epithelial transport of drugs in Caco-2 cell culture. *J Pharm Pharmaceut Sci* **2004,** 7, 186-199.
38. Marrero-Ponce, Y.; Montero-Torres, A.; Zaldivar, C. R.; Veitia, M. I.; Perez, M. M.; Sanchez, R. N. Non-stochastic and stochastic linear indices of the 'molecular pseudograph's atom adjacency matrix': application to 'in silico' studies for the rational discovery of new antimalarial compounds. *Bioorg Med Chem* **2005,** 13, 1293-1304.
39. Marrero-Ponce, Y.; Medina-Marrero, R.; Torrens, F.; Martinez, Y.; Romero-Zaldivar, V.; Castro, E. A. Atom, atom-type, and total non-stochastic and stochastic quadratic fingerprints: a promising approach for modeling of antibacterial activity. *Bioorg Med Chem* **2005,** 13, 2881-2899.
40. Marrero-Ponce, Y.; Marrero, R. M.; Torrens, F.; Martinez, Y.; Bernal, M. G.; Zaldivar, V. R.; Castro, E. A.; Abalo, R. G. 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* **2006,** 12, 255-71.
41. Marrero-Ponce, Y.; Machado-Tugores, Y.; Pereira, D. M.; Escario, J. A.; Barrio, A. G.; Nogal-Ruiz, J. J.; Ochoa, C.; Aran, V. J.; Martinez-Fernandez, A. R.; Sanchez, R. N.; Montero-Torres, A.; Torrens, F.; Meneses-Marcel, A. A computer-based approach to the rational discovery of new trichomonacidal drugs by atom-type linear indices. *Curr Drug Discov Technol* **2005,** 2, 245-65.
42. Marrero-Ponce, Y.; Iyarreta-Veitia, M.; Montero-Torres, A.; Romero-Zaldivar, C.; Brandt, C. A.; Avila, P. E.; Kirchgatter, K.; Machado, Y. 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 Model* **2005,** 45, 1082-100.
43. Marrero-Ponce, Y.; Huesca-Guillen, A.; Ibarra-Velarde, F. 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)* **2005,** 717, 67-79.

44. Marrero-Ponce, Y.; Castillo-Garit, J. A.; Torrens, F.; Romero-Zaldivar, V.; Castro, E. Atom, Atom-Type, and Total Linear Indices of the ¨Molecular Pseudograph´s Atom Adjacency Matrix¨: Application to QSPR/QSAR Studies of Organic Compounds. *Molecules* **2004,** 9, 1100-1123.
45. 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. 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* **2005,** 13, 1005-1020.
46. Marrero-Ponce, Y.; Castillo-Garit, J. A.; Olazabal, E.; Serrano, H. S.; Morales, A.; Castañedo, N.; Ibarra-Velarde, F.; Huesca-Guillen, A.; Jorge, E.; del Valle, A.; Torrens, F.; Castro, E. A. TOMOCOMD-CARDD, a novel approach for computer-aided 'rational' drug design: I. Theoretical and experimental assessment of a promising method for computational screening and in silico design of new anthelmintic compounds. *J. Comput.-Aided Mol. Design* **2004,** 18, 615-634.
47. Marrero-Ponce, Y.; Cabrera, M. A.; Romero-Zaldivar, V.; Bermejo, M.; Siverio, D.; Torrens, F. 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* **2005,** 4 124-150.
48. Marrero-Ponce, Y. Total and local (atom and atom type) molecular quadratic indices: significance interpretation, comparison to other molecular descriptors, and QSPR/QSAR applications. *Bioorg. Med. Chem.* **2004,** 12, 6351-6369.
49. Marrero-Ponce, Y.; Cabrera, M., A.; Romero, V.; Ofori, E.; Montero, L. A. Total and Local Quadratic Indices of the "Molecular Pseudograph's Atom Adjacency Matrix". Application to Prediction of Caco-2 Permeability of Drugs. *Int. J. Mol. Sci.* **2003,** 4, 512-536.
50. Montero-Torres, A.; Vega, M. C.; Marrero-Ponce, Y.; Rolon, M.; Gomez-Barrio, A.; Escario, J. A.; Aran, V. J.; Martinez-Fernandez, A. R.; Meneses-Marcel, A. A novel non-stochastic quadratic fingerprints-based approach for the 'in silico' discovery of new antitrypanosomal compounds. *Bioorg Med Chem* **2005,** 13, 6264-75.
51. Montero-Torres, A.; Garcia-Sanchez, R. N.; Marrero-Ponce, Y.; Machado-Tugores, Y.; Nogal-Ruiz, J. J.; Martinez-Fernandez, A. R.; Aran, V. J.; Ochoa, C.; Meneses-Marcel, A.; Torrens, F. Non-stochastic quadratic fingerprints and LDA-based QSAR models in hit and lead generation through virtual screening: theoretical and experimental assessment of a promising method for the discovery of new antimalarial compounds. *Eur J Med Chem* **2006**.
52. Meneses-Marcel, A.; Marrero-Ponce, Y.; Machado-Tugores, Y.; Montero-Torres, A.; Pereira, D. M.; Escario, J. A.; Nogal-Ruiz, J. J.; Ochoa, C.; Aran, V. J.; Martinez-Fernandez, A. R.; Garcia Sanchez, R. N. A linear discrimination analysis based virtual screening of trichomonacidal lead-like compounds: outcomes of in silico studies supported by experimental results. *Bioorg Med Chem Lett* **2005,** 15, 3838-43.
53. Casanola-Martin, G. M.; Khan, M. T.; Marrero-Ponce, Y.; Ather, A.; Sultankhodzhaev, M. N.; Torrens, F. New tyrosinase inhibitors selected by atomic linear indices-based classification models. *Bioorg Med Chem Lett* **2006,** 16, 324-30.

54. Marrero-Ponce, Y.; Castillo-Garit, J. A. 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. Design* **2005,** 19, 369-83.
55. Marrero-Ponce, Y.; Díaz, H. G.; Romero, V.; Torrens, F.; Castro, E. A. 3D-Chiral quadratic indices of the ¨molecular pseudograph´s atom adjacency matrix¨ and their application to central chirality codification: classification of ACE inhibitors and prediction of r-receptor antagonist activities. *Bioorg. Med. Chem.* **2004,** 12, 5331-5342.
56. Castillo-Garit, J. A.; Marrero-Ponce, Y.; Torrens, F.; Rotondo, R. Atom-Based Stochastic and Non-Stochastic 3D-Chiral Bilinear Indices and Their Applications to Central Chirality Codification. *J. Mol. Graph. Modell.* **2007,** 26, 32-47.
57. Castillo-Garit, J. A.; Marrero-Ponce, Y.; Torrens, F. Atom-Based 3D-Chiral Quadratic Indices. Part 2: Prediction of the Corticosteroid-Binding Globulin Binding Affinity of the 31 Benchmark Steroids Data Set. *Bioorg. Med. Chem.* **2006,** 14, 2398–2408.
58. Marrero-Ponce, Y.; Castillo-Garit, J. A. 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.* **2005,** 19, 369-83.
59. Marrero-Ponce, Y.; Castillo-Garit, J. A.; Castro, E. A.; Torrens, F.; Rotondo, R. 3D-Chiral (2.5) Atom-Based TOMOCOMD-CARDD Descriptors: Theory and QSAR Applications to Central Chirality Codification. *J. Math. Chem.* **2008,** DOI 10.1007/s10910-008-9386-3.
60. Meneses-Marcel, A.; Marrero-Ponce, Y.; Machado-Tugores, Y.; Montero-Torres, A.; Montero Pereira, D.; Escario, J. A.; Nogal-Ruiz, J. J.; Ochoa, C.; Arán, V. J.; Martínez-Fernández, A. R.; García Sánchez, R. N. A Linear Discrimination Analysis Based Virtual Screening of Trichomonacidal Lead-like Compounds: Outcomes of in silico Studies Supported by Experimental Results. *Bioorg. Med. Chem. Lett.* **2005,** 17, 3838-3843.
61. Rivera-Borroto, O. M.; Marrero-Ponce, Y.; Meneses-Marcel, A.; Escario, J. A.; Gómez-Barrio, A.; A., A. V.; Martins-Alho, M. A.; Montero Pereira, D.; Nogal, J. J.; Torrens, F.; Ibarra-Velarde, F.; Vera Montenegro, V.; Huesca-Guille, A.; Rivera, N.; Vogel, V. Discovery of Novel Trichomonacidals Using LDA-Driven QSAR Models and Bond-Based Bilinear Indices as Molecular Descriptors. *QSAR Comb. Sci.* **2008**, DOI: 10.1002/qsar.200610165.
62. Marrero-Ponce, Y.; Castillo-Garit, J. A.; Olazabal, E.; Serrano, H. S.; Morales, A.; Castanedo, N.; Ibarra-Velarde, F.; Huesca-Guillen, A.; Jorge, E.; del Valle, A.; Torrens, F.; Castro, E. A. TOMOCOMD-CARDD, a Novel Approach for Computer Aided 'Rational' Drug Design: I. Theoretical and Experimental Assessment of a Promising Method for Computational Screening and in silico Design of New Anthelmintic Compounds. *J. Comput. Aided Mol. Des.* **2004,** 18, 615-34.
63. 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. 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.* **2005,** 13, 1005-1020.
64. Marrero-Ponce, Y.; Iyarreta-Veitia, M.; Montero-Torres, A.; Romero-Zaldivar, C.; Brandt, C. A.; Avila, P. E.; Kirchgatter, K.; Machado, Y. 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.* **2005,** 45, 1082-100.

65. Montero-Torres, A.; García-Sánchez, R. N.; Marrero-Ponce, Y.; Machado-Tugores, Y.; Nogal-Ruiz, J. J.; Martínez-Fernández, A. R.; Arán, V. J.; Ochoa, C.; Meneses-Marcel, A.; Torrens, F. Non-stochastic quadratic fingerprints and LDA-based QSAR models in hit and lead generation through virtual screening: theoretical and experimental assessment of a promising method for the discovery of new antimalarial compounds. *Eur. J. Med. Chem.* **2006,** 41, 483–493.
66. Vega, M. C.; Montero-Torres, A.; Marrero-Ponce, Y.; Rolón, M.; Gómez-Barrio, A.; Escario, J. A.; Arán, V. J.; Martínez-Fernández, A. R.; Meneses-Marcel, A. New Ligand-Based Approach for the Discovery of Antitrypanosomal Compounds *Bioorg. Med. Chem. Lett.* **2006,** 16, 1898–1904.
67. Montero-Torres, A.; Vega, M. C.; Marrero-Ponce, Y.; Rolón, M.; Gómez-Barrio, A.; Escario, J. A.; Arán, V. J.; Martínez-Fernández, A. R.; Meneses-Marcel, A. A Novel Non-Stochastic Quadratic Fingerprints-Based Approach for the 'in silico' Discovery of New Antitrypanosomal Compounds *Bioorg. Med. Chem.* **2005,** 13, 6264–6275.
68. Arán, V. J.; Asensio, J. L.; Ruiz, J. R.; Stud, M. The Heterocyclization of N',N'-Disubstituted 2-Halobenzohydrazides to 1,1-Disubstituted Indazol-3-Ylioxides. *J. Chem. Res. (S)* **1993,** 218, 1322-1345.
69. Arán, V. J.; Asensio, J. L.; Ruiz, J. R.; Stud, M. *J. Chem. Res. (M)* **1993**, 1322-1345.
70. Arán, V. J.; Asensio, J. L.; Ruiz, J. R.; Stud, M. Reactivity of 1,1-disubstituted Indazol-3-yliooxides: Synthesis of Some Substituted Indazolols and Indazolinones". *J. Chem. Soc., Perkin Trans. 1.* **1993**, 1119-1127.
71. Arán, V. J.; Flores, M.; Muñoz, M.; Ruiz, J. R.; Sánchez-Verdú, P.; Stud, M. *Liebigs Ann.* **1995**, 817-824.
72. Arán, V. J.; Flores, M.; Muñoz, M.; Páez, J. A.; Sánchez-Verdú, P.; Stud, M. *Liebigs Ann.* **1996**, 683-691.
73. Arán, V. J.; Ochoa, C.; Boiani, L.; Buccino, P.; Cerecetto, H.; Gerpe, A.; Gonzalez, M.; Montero, D.; Nogal, J. J.; Gomez-Barrio, A.; Azqueta, A.; Lopez de Cerain, A.; Piro, O. E.; Castellano, E. E. Synthesis and biological properties of new 5-nitroindazole derivatives. *Bioorg. Med. Chem.* **2005,** 13, 3197-207.
74. Ruiz, J. R.; Arán, V. J.; Asensio, J. L.; Flores, M.; Stud, M. Synthesis of Quaternary Indoxyl Derivatives by Intramolecular Cyclization of Some Substituted Acetophenone. *Liebigs Ann. Chem.* **1994**, 679-684.
75. Arán, V. J.; Asensio, J. L.; Molina, J.; Muñoz, P.; Ruiz, J. R.; Stud, M. Approaches to 1,1-disubstituted Cinnolin-3-ylio Oxides: Synthesis and Reactivity of a New Class of Heterocyclic Betaines. *J. Chem. Soc., Perkin Trans.1.* **1997**, 2229-2235.
76. de Castro, S.; Chicharro, R.; Arán, V. J. Synthesis of Quinoxaline Derivatives from Substituted Acetanilides Through Intramolecular Quaternization Reactions. *J. Chem. Soc., Perkin Trans.1.* **2002**, 790-802.
77. Chicharro, R.; de Castro, S.; Reino, J. L.; Arán, V. J. Synthesis of Tri- and Tetracyclic Condensed Quinoxalin-2-ones Fused Across the C-3-N-4 Bond. *Eur. J. Org. Chem.* **2003**, 2314-2326.
78. Marrero-Ponce, Y.; Meneses-Marcel, A.; Machado-Tugores, Y.; Montero Pereira, D.; Escario, J. A.; Nogal-Ruiz, J. J.; Ochoa, C.; Arán, V. J.; Martínez-Fernández, A. R.; García Sánchez, R. N.; Montero-Torres, A.; Torrens., F. A Computer-Based

Approach to the Rational Discovery of New Antitrichomonas Drugs by Atom-Type Linear Indices. *Curr. Drug Disc. Tech.* **2005,** 2, 245-265.
79. Celeste-Vega, M.; Montero-Torres, A.; Marrero-Ponce, Y.; Rolón, M.; Gómez-Barrio, A.; Escario, J. A.; Arán, V. J.; Martínez-Fernández, A. R.; Meneses-Marcel, A. New Ligand-Based Approach for the Discovery of Antitrypanosomal Compounds. . *Bioorg. Med. Chem. Letter.* **2006,** 16, 1898–1904.
80. Patlewicz, G.; Dimitrov, S. D.; Low, L. K.; Kern, P. S.; Dimitrova, G. D.; Comber, M. I. H.; Aptula, A. O.; Phillips, R. D.; Niemela, J.; Madsen, C.; Wedebye, E. V.; Roberts, D. W.; Bailey, P. T.; Mekenyan, O. G. TIMES-SS—A Promising Tool for the Assessment of Skin Sensitization Hazard. A Characterization with Respect to the OECD Validation Principles for (Q)SARs and an External Evaluation for Predictivity. *Regulat. Toxicol. Pharm.* **2007,** 48, 225–239.
81. Marrero-Ponce, Y.; Montero-Torres, A.; Zaldivar, C. R.; Veitia, M. I.; Perez, M. M.; Sanchez, R. N. Non-stochastic and Stochastic Linear Indices of the "Molecular Pseudograph's Atom Adjacency Matrix": Application to 'in silico' Studies for the Rational Discovery of New Antimalarial Compounds. *Bioorg. Med. Chem.* **2005,** 13, 1293-304.
82. OECD. Guidance document on the validation of quantitative structure-activity relationship [QSAR] models. In *OECD Environment Health and Safety Publication, Series on testing and assessment no. 69*, Paris, 2007. http://www.oecd.org/dataoecd/55/35/38130292.pdf.
83. Golbraikh, A.; Tropsha, A. Predictive QSAR Modeling based on Diversity Sampling of Experimental Datasets for the Training and Test Set Selection. *Mol. Divers.* **2002,** 5, 231-243.
84. Negwer, M. *Organic-Chemical Drugs and their Synonyms*. Akademie: Berlin, 1987.
85. The Merck Index. In 12th ed.; Chapman and Hall: 1996.
86. Glasby, J. S. Encyclopedia of Antibiotics. In Woodhouse: Manchester, 1978.
87. Golbraikh, A.; Shen, M.; Xiao, Z.; Xiao, Y. D.; Lee, K. H.; Tropsha, A. Rational Selection of Training and Test Sets for the Development of Validated QSAR Models. *J. Comput. Aided Mol. Des.* **2003,** 17, 241-253.
88. Mc Farland, J. W.; Gans, D. J. Cluster Significance Analysis. In *Chemometric Methods in Molecular Design*, Waterbeemd, H., Ed. VCH Publishers: Weinheim, Ger, 1995; pp 295–307.
89. Marrero-Ponce, Y.; Marrero, R. M.; Torrens, F.; Martinez, Y.; Bernal, M. G.; Zaldivar, V. R.; Castro, E. A.; Abalo, R. G. 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.* **2005**, 1-17.
90. Consonni, V.; Todeschini, R.; Pavan, M. Structure/Response Correlations and Similarity/Diversity Analysis by GETAWAY Descriptors. 1. Theory of the Novel 3D Molecular Descriptors. *J. Chem. Inf. Comput. Sci.* **2002,** 42, 682-692.
91. Gramatica, P. Principles of QSAR Models Validation: Internal and External QSAR. *QSAR Comb. Sci.* **2007,** 26, 694-701.
92. Eriksson, L.; Jaworska, J.; Worth, A. P.; Cronin, M. T.; McDowell, R. M.; Gramatica, P. Methods for Reliability and Uncertainty Assessment and for Applicability Evaluations of Classification- and Regression-Based QSARs. *Environ. Health Perspect.* **2003,** 111, 1361-1375.

93. Atkinson, A. C. *Plots, Transformations and Regression*. Clarendon Press: Oxford, 1985.
94. Papa, E.; Villa, F.; Gramatica, P. Statistically Validated QSARs, based on Theoretical Descriptors, for Modeling Aquatic Toxicity of Organic Chemicals in Pimephales promelas (fathead minnow). *J. Chem. Inf. Model.* **2005,** 45, 1256-1266.
95. Golbraikh, A.; Tropsha, A. Beware of q2! *J Mol Graph Model* **2002,** 20, 269-76.
96. van DAMME, S.; Bultinck, P. Software News and Update A New Computer Program for QSAR-Analysis: ARTE-QSAR. *J. Comput. Chem.* **2007,** 28, 1924-1928.
97. Van Damme, S.; Langenaeker, W.; Bultinck, P. Prediction of blood–brain partitioning: A model based on ab initio calculated quantum chemical descriptors. *J. Mol. Graphics Modell.* **2008,** 26, 1223-1236.
98. Shen, M.; Xiao, Y.; Golbraikh, A.; Gombar, V. K.; Tropsha, A. Development and validation of k-nearest-neighbor QSPR models of metabolic stability of drug candidates. *J. Med. Chem.* **2003,** 46, 3013-20.
99. Kubinyi, H. Chemical Similarity and Biological Activities. *J. Braz. Chem. Soc.* **2002,** 13, 717-726.
100. Eckert, H.; Bajorath, J. Molecular Similarity Analysis in VirtualScreening: Foundations, Limitations and Novel Approaches. *Drug Discov. Today* **2007,** 12, 225-236.
101. Tyler, K. M.; Engman, D. M. The life cycle of Trypanosoma cruzi revisited *Int. J. Parasitol.* **2001,** 31, 472-481.
102. Mondragón, R.; Frixione, E. Ca^{2+} Dependence of conoid extrusion in Toxoplasma gondii Tachyzoites. *J. Eur. Microbiol.* **1996,** 43, 120-127.
103. Patrón, A.; Mondragón, M.; Gonzáles, S.; Ambrosio, J.; Guerrero, A.; Mondragón, R. Identification and purification of actin from the subpellicular network of Toxoplasma gondii tachyzoites. *Int. J. Parasitol.* **2005,** 35, 883-894.
104. Deharo, E.; Garcia, R.; Oporto, P.; Sauvain, M.; Gautret, P.; Ginsburg, H. A non-radiolabeled ferriprotoporphyrin IX biomineralization inhibition test (FBIT) for the high throughput screening of antimalarial compounds. *Exp. Parasitol.* **2002,** 100, 252-256.
105. Desjardins, R. E.; Canfield, C. J.; Haynes, J. D.; Chulay, J. D. Quantitative Assessment of Antimalarial Activity In Vitro by a Semiautomated Microdilution Technique. *Antimicrob. Agents Chemother.* **1979,** 16, 710-718.
106. Chou, K.-C.; Shen, H.-B. Recent progress in protein subcellular location prediction. *Anal. Biochem.* **2007,** 370, 1-16.
107. Van de Waterbeemd, H.; Carter, R. E.; Grassy, G.; Kubinyi, H.; Martin, Y. C.; Tute, M. S.; Willett, P. *Annu. Rep. Med. Chem.* **1998,** 33, 397.
108. Karelson, M. *Molecular Descriptors in QSAR/QSPR*. John Wiley & Sons: New York, 2000.
109. Katritzky, A. R.; Gordeeva, E. V. Traditional topological indices vs electronic, geometrical, and combined molecular descriptors in QSAR/QSPR research. *J. Chem. Inf. Comput. Sci.* **1993,** 33, 835-57.
110. Todeschini, R.; Consonni, V. *Handbook of Molecular Descriptors*. Wiley-VCH: Germany, 2000.
111. *STATISTICA (data analysis software system) vs 6.0*, StatSoft Inc: Tulsa,OK:, 2001.
112. Baldi, A.; Dragonetti, E.; Battista, T.; Groeger, A. M.; Esposito, V.; Baldi, G.; Santini, D. Detection of Circulating Malignant Cells by RT-PCR in Long-Term

Clinically Disease-Free I Stage Melanoma Patients. *Anticancer Res.* **2000,** 20, 3923-3928.
113. Kouznetsov, V. V.; Rivero, C. J.; Ochoa, P. C.; Stashenko, E.; Martínez, J. R.; Montero, P. D.; Nogal, R. J. J.; Fernández, P. C.; Muelas, S. S.; Gómez, B. A.; Bahsas, A.; Amaro, L. J. Synthesis and Antiparasitic Properties of New 4-N-Benzylamino-4-Hetarylbut-1-enes. *Arch. Pharm. (Weinheim, Ger.)* **2005,** 338, 32-37.
114. Kouznetsov, V. V.; Vargas, M. L. Y.; Tibaduiza, B.; Ochoa, C.; Montero, P. D.; Nogal, R. J. J.; Fernández, C.; Muelas, S.; Gómez, A.; Bahsas, A.; Amaro-Luis, J. 4-Aryl(benzyl)amino-4-heteroarylbut-1-enes as Building Blocks in Heterocyclic Synthesis. 4.1 Synthesis of 4, 6-Dimethyl-5-nitro(amino)-2-pyridylquinolines and their Antiparasitic Activities. *Arch. Pharm. (Weinheim, Ger.)* **2004,** 337, 127-132.
115. Vega, M. C.; Rolon, M.; Martinez-Fernandez, A. R.; Escario, J. A.; Gomez-Barrio, A. A New Pharmacological Screening Assay with Trypanosoma cruzi Epimastigotes Expressing Beta-Galactosidase. *Parasitol. Res.* **2005,** 95, 296-298.
116. Hattori, Y.; Nakanishi, N. Effects of Cyclosporin A and FK506 on Nitric Oxide and Tetrahydrobiopterin Sysnthesis in Bacterial Lipopolysaccharide-Treated J774 Macrophages. *Cell. Immunol.* **1995,** 165, 7-11.

ANNEXES

Table 1. Prediction Performances and Statistical Parameters for LDA-based QSAR Models in the Training Set.

Eqs.	Atomic Labels[a]	Matthews corr. coeff.	Accuracy 'Q_{total}' (%)	Specificity (%)	Sensitivity 'hit rate' (%)	False '+' rate (%)	Landa Wilks	D^2	F
				Non-Stochastic Linear Indices					
1	(M)	0.72	86.71	83.41	83.82	11.33	0.49	4.28	73.36
2	(P)	0.80	90.48	91.49	84.31	5.33	0.49	4.29	73.66
3	(V)	0.79	89.88	90.05	84.31	6.33	0.493	4.24	72.75
4	(K)	0.80	90.48	91.05	84.80	5.667	0.467	4.72	80.83
				General model (combining all atomic labels)					
5	(NS)	0.82	91.27	92.11	85.78	5.00	0.467	4.98	80.83
				Stochastic Linear Indices					
6	(M)	0.70	85.71	84.02	79.90	10.33	0.60	2.73	40.83
7	(P)	0.76	88.49	88.83	81.86	7.00	0.52	3.77	64.63
8	(V)	0.79	89.68	88.38	85.78	7.67	0.52	3.74	56.02
9	(K)	0.76	88.29	87.96	82.35	7.667	0.51	3.92	52.06
				General model (combining all atomic labels)					
10	(SS)	0.82	91.27	93.48	84.31	4.00	0.46	4.83	96.93
				Mixing all MDs (non-stochastic and stochastic indices)					
11	**(NS-SS)**	**0.86**	**93.06**	**92.89**	**89.71**	**4.67**	**0.435**	**5.35**	**107.3**

[a]M: atomic mass, P: atomic polarizability, K: atomic Mullinken electronegativity, V: van der Waals atomic volume.[90] NS, SS and NS-SS jeans non-stochastic MDs, stochastic MDs and whole set of MDs, respectively.

Table 2. Prediction Performances for LDA-based QSAR Models in the Test Set.

Eqs.	Atomic Labels[a]	Matthews corr. coeff.	Accuracy 'Q_{total}' (%)	Specificity (%)	Sensitivity 'hit rate' (%)	False '+' rate (%)
			Non-Stochastic Linear Indices (NS)			
1	(M)	0.66	85.23	70.69	82.00	13.49
2	(P)	0.61	83.52	68.42	78.00	14.29
3	(V)	0.61	84.09	71.15	74.00	11.90
4	(K)	0.66	84.09	66.67	88.00	17.46
			General model (combining all atomic labels)			
5	**(NS)**	0.67	85.80	71.19	84.00	13.49
			Stochastic Linear Indices (S)			
6	(M)	0.35	71.02	49.23	64.00	26.19
7	(P)	0.41	74.43	54.24	64.00	21.43
8	(V)	0.57	81.25	63.93	78.00	17.46
9	(K)	0.42	73.30	52.17	72.00	26.19
			General model (combining all atomic labels)			
10	**(SS)**	0.52	79.55	62.07	72.00	17.46
			Mixing all MDs (non-stochastic and stochastic indices, NS-S)			
11	(NS-SS)	**0.81**	**92.05**	**83.33**	**90.00**	**7.14**

[a]M: atomic mass, P: atomic polarizability, K: atomic Mullinken electronegativity, V: van der Waals atomic volume.[90] NS, SS and NS-SS jeans non-stochastic MDs, stochastic MDs and whole set of MDs, respectively.

Table 3. Results of Ligand-based *in silico* Screening by Using C_I and C_E.

Compound*	*Result by using whole set of* C_I											C_E Class[b]
	ΔP%[a] Eq. 1	ΔP%[a] Eq. 2	ΔP%[a] Eq. 3	ΔP%[a] Eq. 4	ΔP%[a] Eq. 5	ΔP%[a] Eq. 6	ΔP%[a] Eq. 7	ΔP%[a] Eq. 8	ΔP%[a] Eq. 9	ΔP%[a] Eq. 10	ΔP%[a] Eq. 11	
9	88.52	65.68	71.94	72.28	86.99	44.53	88.64	95.44	90.49	96.78	92.95	11
10	82.37	23.57	81.68	71.39	86.91	41.86	67.79	96.88	87.79	97.56	91.73	11
11	82.77	20.49	81.90	71.52	87.77	34.61	67.80	97.12	89.02	97.52	91.84	11
12	83.36	15.96	81.87	75.16	88.81	37.49	68.75	97.40	90.31	97.53	91.82	11
13	74.25	63.10	63.94	75.34	78.29	56.37	60.75	98.39	68.54	98.26	93.50	11
14	77.56	70.57	58.60	86.10	77.77	68.55	80.79	98.51	72.14	98.29	95.17	11
15	79.64	24.10	83.89	77.29	84.65	53.17	58.56	96.29	84.39	97.79	89.31	11
16	80.10	21.02	84.09	77.39	85.63	46.80	58.56	96.58	85.94	97.76	89.46	11
17	80.77	16.50	84.06	80.37	86.84	49.35	59.72	96.90	87.56	97.76	89.43	11
18	70.47	63.42	67.92	80.52	74.73	65.54	50.10	98.08	60.90	98.43	91.58	11

*The molecular structures of the compounds represented with codes (numbers) are shown in Scheme. [a]ΔP% = [P(Active) - P(Inactive)]*x*100 of each compounds in this screening set (see experimental section). Classification of each compounds using every obtained C_I models in the following order: Eq. 1-11. Here, in order to consider every query molecule as active chemical we used ΔP%>15%, because with this cut-off we avoid the not classified example as well as the risk of false active can be less. Classification of each compounds using the C_e (see Eq. 13-17 in Experimental Section).

Table 4. Percentages of Citostatic and/or Citocidal Activity [brackets] for the Three Concentrations Assayed *in vitro* Against *Trichomonas vaginalis.*

Compound*	Obs [a]	*in vitro* activity (μg/mL)[b]					
		$\%CA_{24h}$ [$\%C_{24h}$]			$\%CA_{48h}$ [$\%C_{48h}$]		
		100	**10**	**1**	**100**	**10**	**1**
9	-	29.39	11.43	1.22	28.33	14.68	0
10	-	75,61	21,02	3,53	34,26	1,64	0
11	+	**[99,37]**	20,94	0	**[100]**	5,74	0
12	+	**[100]**	12,94	2,35	**[100]**	0	0
13	+	**[100]**	83,76	3,53	**[100]**	44,06	0
14	+	**[100]**	45.71	8.98	**[100]**	11.26	0
15	++	**[100]**	**[89,25]**	0	**[100]**	67,7	4,1
16	++	**[100]**	**[92,63]**	0	**[100]**	86,52	0
17	++	**[100]**	**[91,61]**	10,98	**[100]**	70,41	2,87
18	+	**[100]**	70,98	4,71	**[100]**	23,28	4,1
Metronidazole	+++	**[100]**	**[99,1]**	**[98,0]**	**[100]**	**[100]**	**[99,5]**

*The molecular structures of the compounds represented with codes (numbers) are shown in Scheme. [a]Observed (experimental activity) classification against *T. vaginalis*. [b]Pharmacological activity of each tested compound, which as added to the cultures at doses of 100, 10 and 1μg/mL: $\%CA_{\#}$ = Cytostatic activity$_{(24 \text{ or } 48 \text{ hours})}$ and **[$\%C_{\#}$]** = Cytocidal activity(% of reduction)$_{(24 \text{ or } 48 \text{ hours})}$. Metronidazole was used as positive control (concentrations for metronidazole were 2, 1 and 0.5 mg/mL, respectively).

Table 5. Antitrypanosomal Activity and Inespecific Citotoxicity at Three Different Concentrations (100, 10 and 1µg/mL) Assayed *In Vitro* against *Tripanosoma cruzi* and Macrophagic Cells, Respectively.

Compound*	Obs.[a]	Concentration (µg/mL)	% Anti-epimastigotes[b] ± % SD	% Cytotoxicity[c] ± % SD
9	NT	100 10 1	NT	NT
10	+	100 10 1	83,54 ± 0,44 5,35 ± 0,25 4,38 ± 0,30	0 ± 0,55 0 ± 2,19 0 ± 2,14
11	+	100 10 1	82,4 ± 0,68 17,68 ± 1,24 1,78 ± 8,63	3,36 ± 1,47 0 ± 1,51 0 ± 1,97
12	+	100 10 1	97,73 ± 0,45 23,84 ± 1,27 8,35 ± 5,11	59,14 ± 1,77 5,78 ± 0,58 0 ± 1,07
13	+	100 10 1	87,83 ± 0,06 56,77 ± 1,41 12,49 ± 1,85	100 ± 0,15 13,25 ± 0,46 9,89 ± 1,21
14	NT	100 10 1	NT	NT
15	-	100 10 1	6,36 ± 4,81 2,51 ± 5,97 0 ± 3,38	49,25 ± 0,4 0 ± 2,26 0 ± 1,25
16	+	100 10 1	79,12 ± 3,86 60,68 ± 2,78 7,93 ± 4,42	61,38 ± 0,53 11,57 ± 2,01 NT
17	-	100 10 1	65,46 ± 5,47 15,38 ± 2,83 0 ± 3,84	75,75 ± 0,9 20,24 ± 1,2 N
18	-	100 10 1	19,78 ± 5,94 15,62 ± 5,06 11,77 ± 4,35	99,44 ± 0,2 24,44 ± 0,26 NT
Nifurtimox	+	100 10 1	98.73 ± 0.5 90.0 ± 1.8 75.5 ± 3.9	25.9 ± 3.9 0.6 ± 3.9 0.0 ± 2.1

*The molecular structures of the compounds represented with codes (numbers) are shown in Scheme. [a]Observed (experimental activity) classification against *T. cruzi*. [a]Experimentally observed activity (compounds with %Anti-epimastigote>70 at 100 (µg/mL) were considered as active ones), [b]Anti-epimastigotes percentage and ±standard deviation (SD). [c]Inespecific citotoxicity in macrophages cells and standard deviation (SD). NT means not tested. Reference drug and positive control: Nifurtimox.

Table 6. Efficacy against *Toxoplasma gondii* Tachyzoites.

Compound*	Obs.[a]	% Tachyzoites Parasites[b]			
		1mM	500µM	200µM	100µM
9	-	73	93	96	95
10	+	0	0	85	91
11	+	0	0	68	92
12	+	0	0	82	88
13	NT	NT	NT	NT	NT
14	-	81	93	90	81
15	-	38	58	84	89
16	-	36	40	90	96
17	±	0	71	77	81
18	NT	NT	NT	NT	NT
DMSO	-	85	93	92	91

*The molecular structures of the compounds represented with codes (numbers) are shown in Scheme. [a]Observed (experimental activity) against *Toxoplasma gondii* Tachyzoites (RH strain). [b]Biochemical studies of percentages of parasites (tachyzoites) for every chemicals evaluated in the range of 1mM, 500µM, 200µM, 100µM. DMSO: Dimethyl sulfoxide.

Table 7. *In Vitro* Antimalarial Activity likes Function of Ferriprotoporphyrin IX Biocrystallization Inhibition Test and Radioisotopic Microtest in *strain* 3D7 of *Plasmodium falciparum*.

Compound*	**Obs.[a]**	Ferriprotoporphyrin IX biocrystallization inh. test	Radioisotopic microtest in strain 3D7 of *Plasmodium falciparum*
		IC_{50} [mg/mL][b]	IC_{50} [mg/mL][c]
9 VAM2-9	-	> 2	> 10
10 VAM2-10	-	> 2	> 10
11 VAM2-11	-	> 2	> 10
12 VAM2-12	+	> 2	**5,72**
13 VAM2-13	+	**1,53**	> 10
14 VAM2-14	-	> 2	> 10
15 VAM2-15	-	> 2	> 10
16 VAM2-16	-	> 2	> 10
17 VAM2-17	+	**1,95**	> 10
18 VAM2-18	++	**0,95**	**6,47**
Chloroquine	++	**0.04**	**0,04**

*The molecular structures of the compounds represented with codes (numbers) are shown in Scheme. [a]Observed (experimental activity) likes function of two diferent *in vitro* assays. [b]IC_{50} values calculate from the percentage of inhibition obtained in ferriprotoporphyrin IX biocrystallization inhibition test (IC_{50}>2 µg/mL were considered as inactives). [c]IC_{50} values calculate from the percentage of inhibition obtained in radioisotopic microtest in strain 3D7 of *Plasmodium falciparum* (IC_{50}>10 µg/mL were considered as inactives). Chloroquine was used as antimalarial reference drug in both assays.

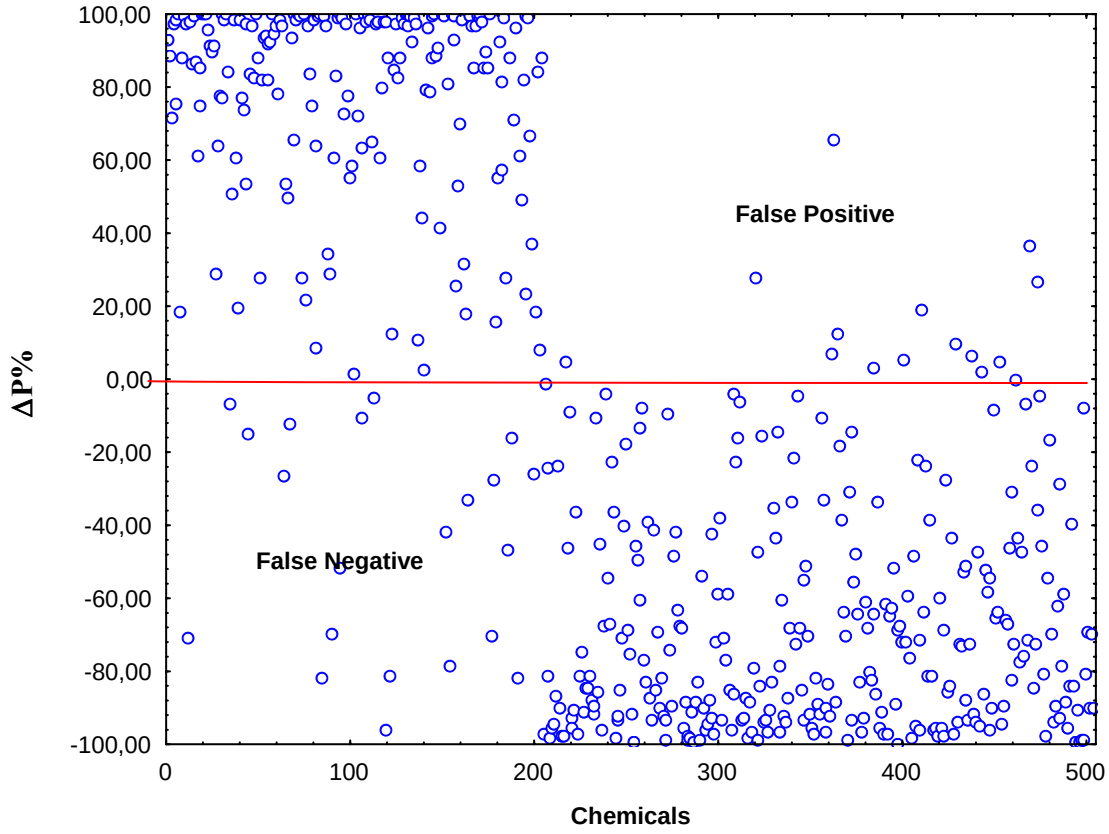


Figure 1. Plot of the ΔP% from Eq. **11** for every compound in the training set. Compounds 1-204 and 205-504 are active and inactive, respectively.

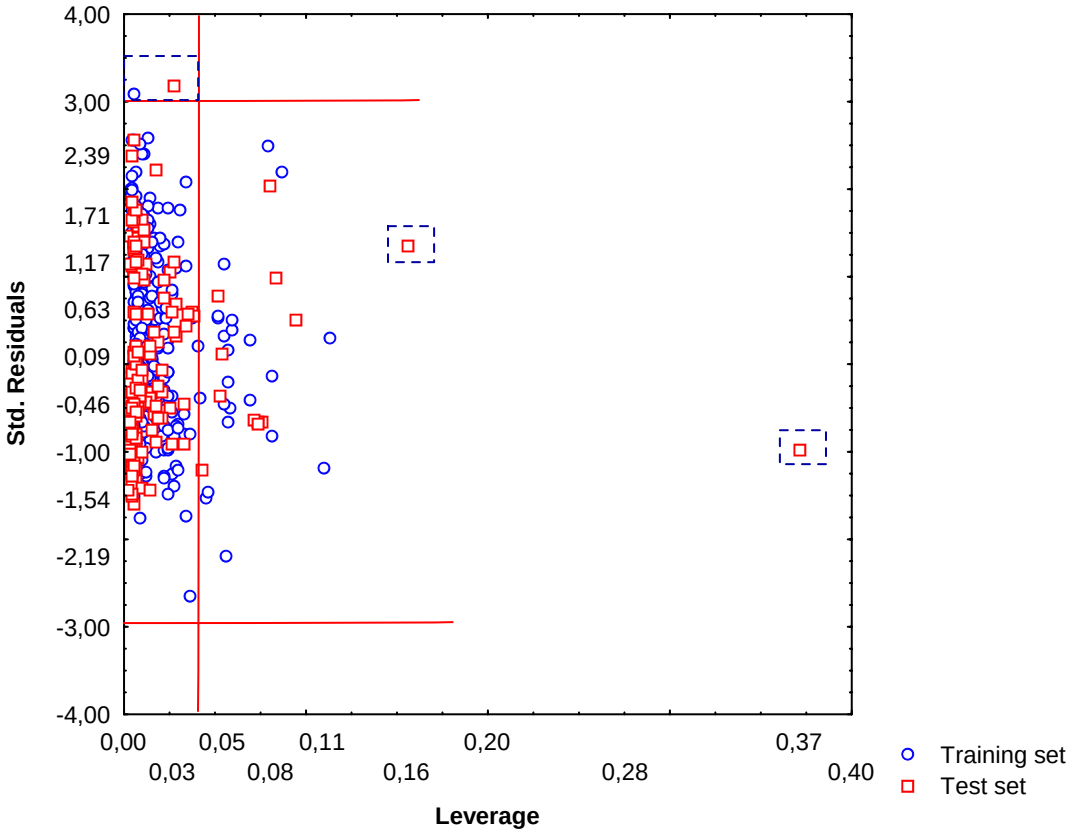


Figure 2. William plot of Eq. **11**: outlier will be chemicals are points with standardized residuals greater than three standard deviation units; influential chemicals are points with high leverage values higher than the *threshold* or *cut-off* value $h^* = 0.042$. The training and test sets are represented by blues circles and red squares, respectively.

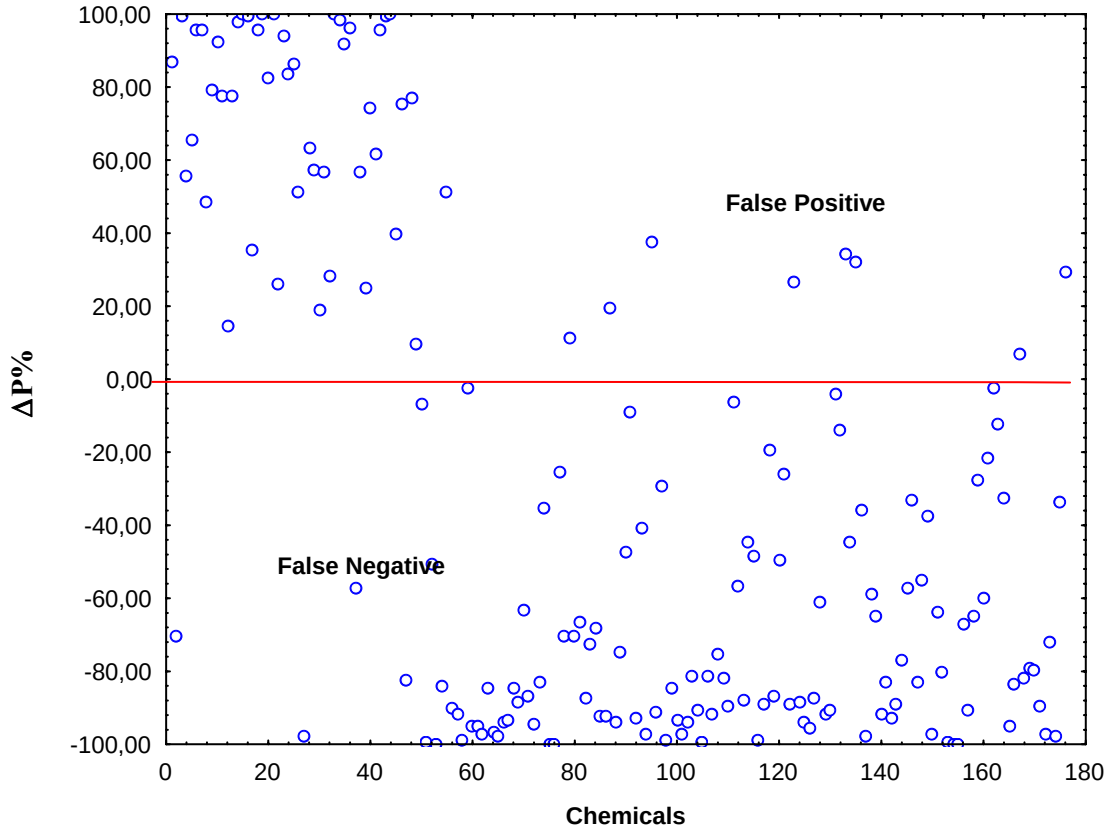


Figure 3. Plot of the ΔP% from Eq. **11** for every compound in the test set. Compounds 1-50 and 51-176 are active and inactive, respectively.

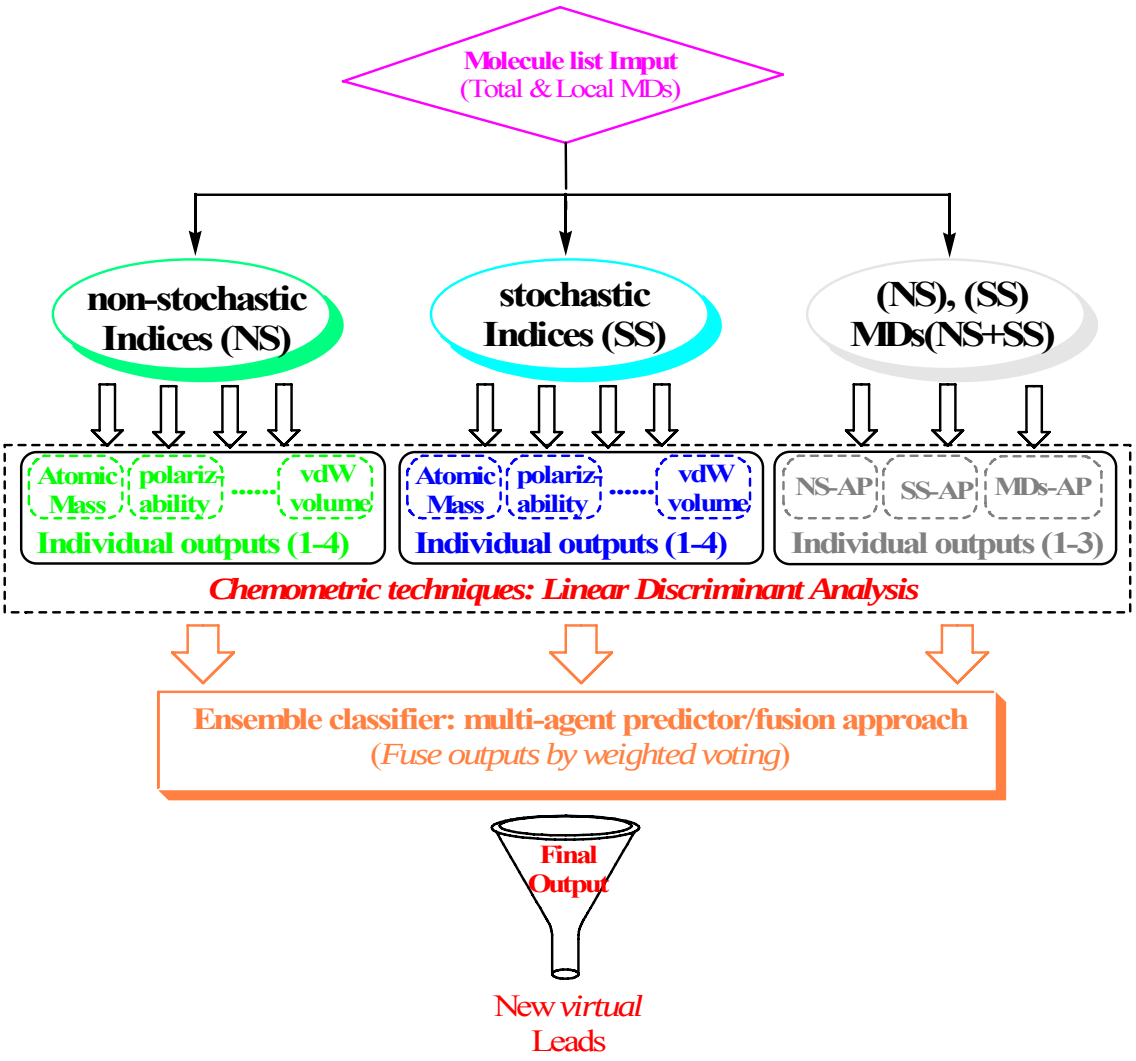


Figure 4. Flowchart illustrating how the individual classifiers are fused into the ensemble classifier through a voting system. Here we show the fuse the discriminant functions by using ***TOMOCOMD-CARDD*** MDs into a prediction engine.

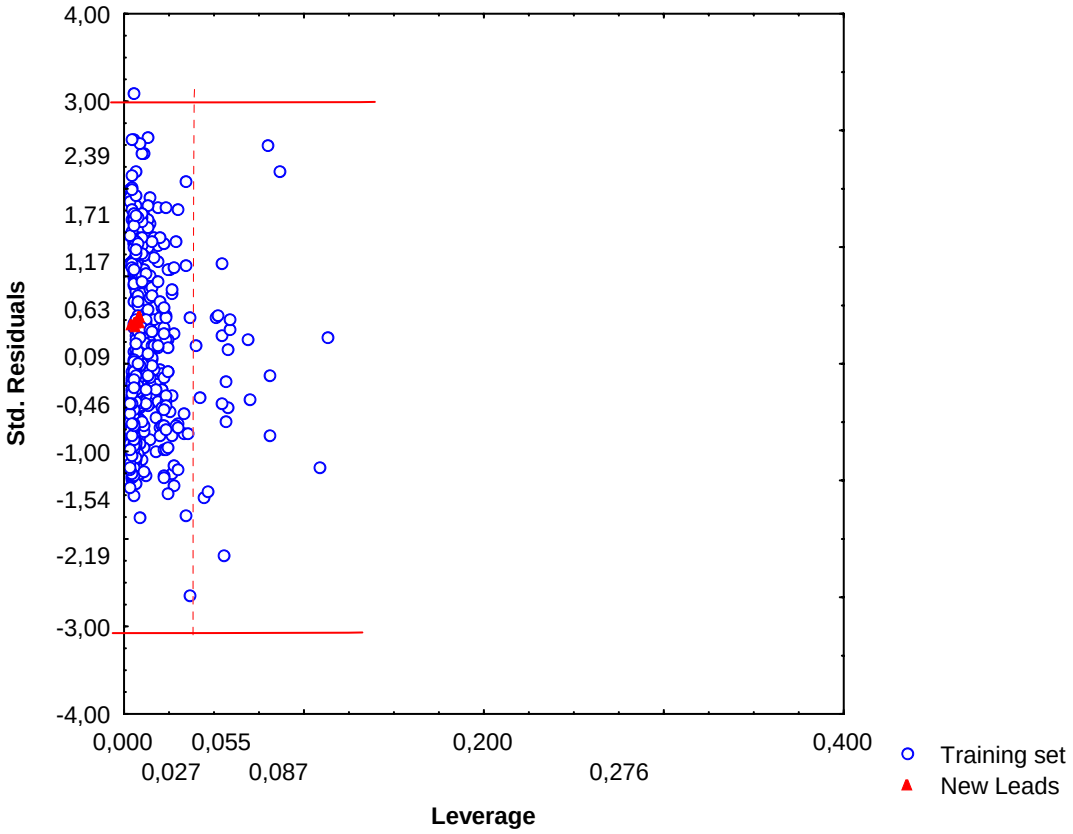
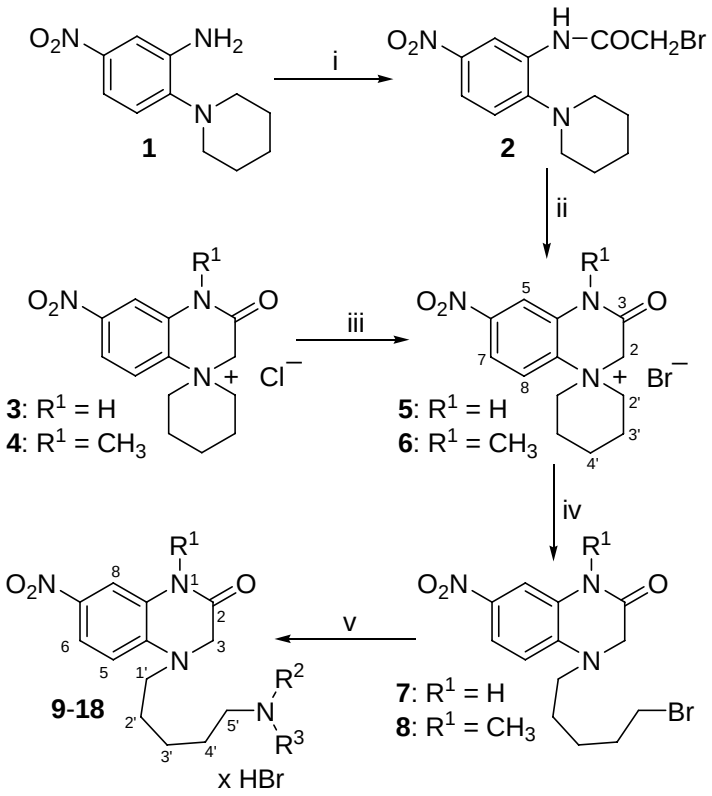


Figure 5. LDA models applicability domain for learning and new leads series. The training is represented by blues circles and the new compounds are represented by red triangles.

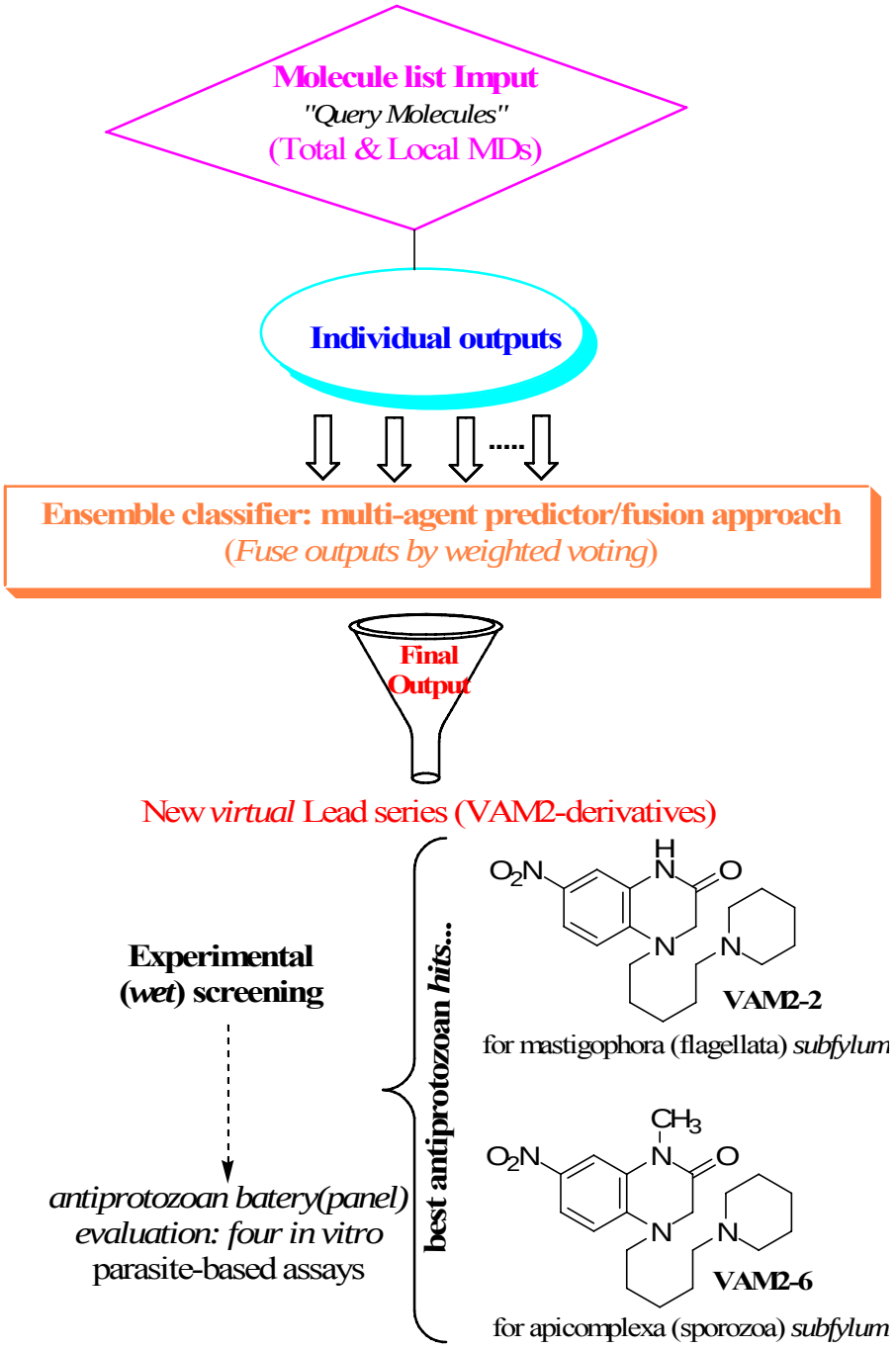
Scheme[a]



R^1 = H	R^1 = CH_3	R^2,R^3
9	**14**	CH_3,CH_3
10	**15**	$[CH_2]_4$
11	**16**	$[CH_2]_5$
12	**17**	$[CH_2]_6$
13	**18**	*o*-CH_2-C_6H_4-$[CH_2]_2$

[a]Reagents and conditions: (i) $BrCH_2COBr$, acetone, r. t., 30 min. (ii) CH_3NO_2, reflux, 25 min. (iii) 48% aq. HBr, vacuum evaporation to dryness (3 times) (iv) CH_3NO_2, reflux (48 h for **7** and 24 h for **8**), argon (v) R^2R^3NH, dioxane, 100-110 ºC (autoclave) or reflux, 5-10 h.

195 **TABLE OF CONTENTS GRAPHIC**



196
197
198 Flowchart illustrating how the individual classifiers are fused into the ensemble
199 classifier through a voting system in order to discovery new antiprotozoan *hits* and
200 *leads*
201