

[c005]

# Bond-Based Quadratic *TOMOCOMD-CARDD* Molecular Indices & Statistical Techniques for New Antitrichomonal Drug-like Compounds Discovery

Alfredo Meneses-Marcel,<sup>a,b</sup> Oscar M. Rivera-Borroto,<sup>a,c</sup> Yovani Marrero-Ponce,<sup>a,d\*</sup> Alina Montero,<sup>a</sup> Yanetsy Machado Tugores,<sup>a</sup> José Antonio Escario,<sup>b</sup> Alicia Gómez Barrio,<sup>b</sup> David Montero Pereira,<sup>b</sup> Juann José Nogal,<sup>b</sup> Vladimir.V. Kouznetsov,<sup>c</sup> Cristian Ochoa Puentes,<sup>c</sup> Arnold R. Bohórquez,<sup>c</sup> Ricardo Grau,<sup>c</sup> Nilo Castañedo Cancio,<sup>a</sup> Francisco Torrens,<sup>d</sup> Froylán Ibarra-Velarde,<sup>f</sup> Richard Rotondo,<sup>g</sup> Ysaías J. Alvarado,<sup>h</sup> Christian Vogel,<sup>i</sup> Lizet Rodríguez-Machin,<sup>a</sup>

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

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

<sup>c</sup>Bioinformatics Group, Center of Studies on Bioinformatics (CEI), Faculty of Mathematics, Physics and Computer Science. Central University of Las Villas, Santa Clara, 54830, Villa Clara, Cuba.

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

<sup>e</sup>Laboratorio de Química Orgánica y Biomolecular. Escuela de Química. UIS. Bucaramanga. Colombia.

<sup>f</sup>Department of Parasitology, Faculty of Veterinarian Medicinal and Zootecnic, UNAM, Mexico, D.F. 04510, Mexico

<sup>g</sup>Mediscovery, Inc. Suite 1050, 601 Carlson Parkway, Minnetonka, MN 55305, USA.

<sup>h</sup>Laboratorio de Electrónica Molecular, Departamento de Química, Modulo II, grano de Oro, Facultad Experimental de Ciencias, La Universidad del Zulia (LUZ), Venezuela.

<sup>i</sup>Universität Rostock, Institut für Chemie, Abteilung für Organische Chemie, Albert-Einstein-Straße 3a, 18059 Rostock.

\*Corresponding author.



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



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



**e-mail:** [ymarrero77@yahoo.es](mailto:ymarrero77@yahoo.es); [yovani.marrero@uv.es](mailto:yovani.marrero@uv.es); [ymponce@gmail.com](mailto:ymponce@gmail.com)  
or [yovanimp@qf.uclv.edu.cu](mailto:yovanimp@qf.uclv.edu.cu)



**URL:** <http://www.uv.es/yoma/>

**Running head:** *New Antitrichomona Drug-like Chemicals*

**Abstract:** New antitrichomonal agents are needed to combat emerging metronidazole-resistant trichomoniasis and reduce the side-effects associated with currently available drugs. Toward this end, bond-based quadratic indices, new *TOMOCOMD-CARDD* molecular descriptors, and linear discriminant analysis (LDA) were used to discover novel, potent, and non-toxic lead trichomonacidal chemicals. Two discriminant functions were obtained with the use of non-stochastic and stochastic total and bond-type quadratic indices for heteroatoms. The obtained LDA-based QSAR models, using non-stochastic and stochastic indices, were able to classify correctly 87.91% (87.50%) and 89.01% (84.38%) of the chemicals in training (test) sets, respectively. They showed large Matthews' correlation coefficients (*C*) of 0.75 (0.71) and 0.78 (0.65) for the training (test) sets, correspondingly. The result of predictions on the 10% *full-out* cross-validation test also evidenced the robustness of the obtained models. Later, both models were applied to the *virtual* screening of 12 compounds already proved against *Trichomonas Vaginalis* (Tv). As a result, they correctly classified 10 out of 12 (83.33%) and 9 out of 12 (75.00%) of the chemicals, respectively; which is a more important criterion for validating the models. In addition, these classification functions were also applied to a library of twenty-one chemicals in order to find new lead antitrichomonal agents. These compounds were synthesized and tested for *in vitro* activity against Tv. As expected, theoretical results almost coincided with experimental ones since there was obtained a correct classification for both models of 95.24% (20 out of 21) of the chemicals. Out of the twenty-one compounds that were screened, and synthesized, two molecules (chemicals G-1, UC-245), showed high to moderate cytotoxic activity at the concentration of 10µg/ml, other two compounds (G-0 and CRIS-148) showed high cytotoxic activity only at the concentration of 100µg/ml, and the remaining chemicals (from CRIS-105 to CRIS-153 except CRIS-148) were inactive at these assayed concentrations. Finally, the best candidate, G-1 (cytotoxic activity of 100% at 10µg/ml) was *in vivo* assayed in ovariectomized Wistar rats achieving promissory results as a trichomonacidal drug-like compound. The LDA-based QSAR models presented here can be considered as a computer-assisted system that could potentially significantly reduce the number of synthesized and tested compounds and increase the chance of finding new chemical entities with antitrichomonal activity.

**Keywords:** *TOMOCOMD-CARDD* Software, Bond-based Quadratic Indices, LDA-based QSAR Model, Virtual Screening, Lead Generation, Trichomonacidal, Cytostatic and Cytotoxic Activity.

*“Models are to be used, not believed”*

Menger, F.M. *J. Am. Chem. Soc.* 107 (1985) 3105.

## 1. INTRODUCTION

*Trichomonas vaginalis* (Tv) is the causative agent of the most common, non-viral, sexually transmitted disease. An estimated 3 million American women (approximately 180 million women worldwide) are infected with Tv every year [1]. This parasite is the main cause of vaginitis, cervicitis and urethritis in women and may be responsible for prostatitis and other genito-urinary syndromes in men [2,3]. Infection with this organism has been linked to various additional pathologic manifestations, including cervical neoplasia [4-7], atypical pelvic inflammatory disease [8], and tubal infertility [9], and has been reported to be a risk factor in the development of posthysterectomy cuff cellulitis [10]. Infection with Tv has also been related to premature rupture of placental membranes, and low birth weight [11,12]. Intrauterine transmission of cytomegalovirus has been reported to be increased by Tv infection [13]. As similar, this infection can elevate the risk of acquiring human immunodeficiency virus [14, 15].

The introduction of nitroheterocyclic drugs in the late 1950s and the 1960s heralded a new era in the treatment of infections caused by gram-negative and -positive bacteria and a range of pathogenic protozoan parasites. The antibiotic azomycin (a 2-nitroimidazole), isolated in Japan from a streptomycete, was the first active nitroimidazole to be discovered [16], and acted as the main impetus for the systematic search for drugs with activity against anaerobic protozoa. This led to the synthesis of the 5-nitroimidazole, *metronidazole* (1-b-hydroxyethyl-2-methyl-5-nitroimidazole) and the demonstration of its activity against Tv [17].

Since then, metronidazole (MTZ) has been the drug of choice for treating trichomoniasis and is currently the only drug licensed for this purpose in the United States. The recommended MTZ regimen results in cure rates of approximately 95% [18]. In fact, MTZ is the drug now most widely used in the treatment of anaerobic protozoan parasitic infections caused by *Tv*, *Giardia duodenalis*, and *Entamoeba histolytica* [17,19-22]. In addition, it is remarkably safe compared to the most toxic antiprotozoal products [23].

MTZ enters the cell through diffusion [24] and is activated in the hydrogenosomes of *Tv* [25]. Here, the nitro group of the drug is anaerobically reduced by pyruvate-ferredoxin oxidoreductase [25]. This results in cytotoxic nitro radical-ion intermediates that break the DNA strands [26]. The response is rapid: cell division and motility cease within 1h and cell death occurs within 8h as seen in cell culture [27].

Although there are clinical reports [28-35] that document the refractoriness of infections with *Tv* to treatment with MTZ, susceptibility tests have failed to demonstrate conclusively that the parasites isolated from such cases after treatment were resistant to this drug [36,37]. Thus, the resistance of *Tv* has not been generally accepted as the factor responsible for failure of MTZ therapy [38], since reinfection, irregular medication, poor absorption of the drug, and its inactivation by the vaginal flora have not been excluded.[37,39,40]. However, a strain of *Tv*, unequivocally resistant to MTZ, was recently isolated from a female patient who had not responded to two courses of treatment with this agent. The current report is concerned with the isolation of this strain and its *in vitro* and *in vivo* susceptibilities to MTZ and other 5-nitroimidazole derivatives [41].

Although MTZ resistance has been considered rare, treatment of these rare patients who do not respond to treatment is extremely problematic for physicians and is associated with enormous patient suffering [42]. A good alternative to palliate this problem could be clinical treatment with other nitroimidazoles but unfortunately all of them have similar modes of antibacterial activity to MTZ [43], and so resistance to MTZ often includes resistance to the other nitroimidazoles [44].

Currently, it is clear that new trichomonocidal agents are needed to treat resistant organisms. However, the great cost associated to the development of new compounds and the small economic size of the market for antiprotozoal drugs makes this development slow. For this reason, it is necessary to develop computational methods permitting theoretical *–in silico–* evaluations of trichomonocidal activity for virtual libraries of chemicals before these compounds are synthesized in the laboratory. This ‘*in silico*’ world of data, analysis, hypothesis, and models that reside inside a computer is alternative to the ‘real’ world of synthesis and screening of compounds in the laboratory [45,46].

At present, many large pharmaceutical industries have reoriented their research strategies seeking to solve the problem of generation/selection of novel chemical entities (NCEs), one of the major bottlenecks in the drug discovery pipeline. In fact, currently most integration projects include efforts to integrate the data associated with NCE generation [47]. Alternatively, several approaches to the computer-aided molecular design and high-throughput *in silico* screening (or virtual high-throughput screening) have been introduced in the literature [48]. Nevertheless, novel computational methods and strategies are required to deliver a system that significantly reduces the time-to-

market and research and development (R&D) spendings, and increase the rate at which NCEs progress through the pipeline. Such studies if they are implemented successfully can deliver substantial benefits and act as the bedrock for NCE selection [47].

In this context, our research group has recently introduced a novel scheme to perform rational *-in silico-* molecular design (or selection/identification of lead drug-like chemicals) and QSAR/QSPR studies, known as ***TOMOCOMD-CARDD*** (acronym of ***Topological MOlecular COMputer Design-Computer Aided “Rational” Drug Design***) [49]. This method has been developed to generate 2D (topologic), 2.5 (3D-chiral) and 3D (topographic and geometric) molecular descriptors based on the application of the discrete mathematics and linear algebra theory to chemistry. In this sense, atomic, atom-type, atom-group and total linear, bilinear and quadratic molecular fingerprints have been defined in analogy to the linear, bilinear and quadratic mathematical maps [50,51]. This *in silico* method has been successfully applied to the prediction of several physical, physicochemical and chemical properties of organic compounds [50-53]. In addition, ***TOMOCOMD-CARDD*** has been extended to consider three-dimensional features of small/medium-sized molecules based on the trigonometric-3D-chirality-correction factor approach [54]. This strategy has also been useful for the prediction of the pharmacokinetic properties of organic compounds [55-57], and the selection of novel subsystems of compounds having a desired property/activity [58-63].

Later, promising results have been found in the modeling of the interaction between drugs and HIV-1 RNA packaging region in the field of bioinformatics using the ***TOMOCOMD-CANAR*** (***Computed-Aided Nucleic Acid Research***) approach [64,65].

Finally, an alternative formulation of our approach for structural characterization of proteins was carried out recently [66,67]. This extended method [***TOMOCOMD-CAMPS*** (***Computed-Aided Modeling in Protein Science***)] was used to encompass protein stability studies by means of a combination of protein linear or quadratic indices (macromolecular fingerprints) and statistical (linear and nonlinear model) methods [66,67].

Recently, some of present authors have proposed a new extended local (bond and bond-type) and total (whole) molecular descriptors based on the adjacency of edges and based on quadratic maps similar to those typically defined by mathematicians in linear algebra. These researchers also proposed a new matrix representation of the molecule on the “stochastic” adjacency of edges and quadratic indices derived from there. These descriptors, called bond-based quadratic indices, encode topological information given by the molecular graph, weighted by chemical information encoded in selected bond weightings. Finally, the correlation ability of the new descriptors is tested in a QSPR and QSAR studies [68].

The main objective of this work was to use non-stochastic and stochastic bond-type quadratic indices to generate predictive LDA (linear discriminant analysis)-based QSAR models enabling the selection of new hits and lead drug-like compounds with antitrichomonal activity. The *in vitro* and *in vivo* evaluation of a new lead series of heterocyclic compounds with antitrichomonal activity is also presented.

## **2. METHODOLOGY**

### **2.1. *TOMOCOMD-CARDD* Approach and 2D Bond-Based Quadratic Indices.**

**TOMOCOMD** is an interactive program for molecular design and bioinformatic research [49]. It is composed of four subprograms; each one of them allows drawing the structures (drawing mode) and calculating molecular 2D/3D (calculation mode) descriptors. The modules are named **CARDD** (Computed-Aided ‘Rational’ Drug Design), **CAMPS** (Computed-Aided Modeling in Protein Science), **CANAR** (Computed-Aided Nucleic Acid Research) and **CABPD** (Computed-Aided Bio-Polymers Docking). In the present report, we outline salient features concerned with only one of these subprograms, **CARDD** and with the calculation of non-stochastic and stochastic 2D bond-based quadratic indices.

**2.1.1. Theoretical Scaffold.** The basis of the extension of quadratic indices that will be given here is the edge-adjacency matrix considered and explicitly defined in the chemical graph-theory literature [69,70], and rediscovered by Estrada as an important source of new molecular descriptors [71-76]. In this section, we first will define the nomenclature to be used in this work, then the atom-based molecular vector ( $\bar{x}$ ) will be redefined for bond characterization using the same approach as previously reported, and finally some new definition of bond-based non-stochastic and stochastic quadratic indices with its peculiar mathematical properties will be given.

**2.1.1.1. Background in Edge-Adjacency Matrix and New Edge-Relations: Stochastic Edge-Adjacency Matrix.** Let  $G = (V, E)$  be a simple graph, with  $V = \{v_1, v_2, \dots, v_n\}$  and  $E = \{e_1, e_2, \dots, e_m\}$  being the vertex- and edge-sets of  $G$ , respectively. Then  $G$  represents a molecular graph having  $n$  vertices and  $m$  edge (bonds). The edge-adjacency matrix  $E$  of  $G$  (likewise called bond adjacency matrix,  $B$ ) is a square and symmetric matrix whose elements  $e_{ij}$  are 1 if and only if edge  $i$  is adjacent to edge  $j$  [71,74,77]. Two edges are



adjacent if they are incidental to a common vertex. This matrix corresponds to the vertex-adjacency matrix of the associated line graph. Finally, the sum of the  $i$ th row (or column) of  $E$  is named the edge degree of bond  $i$ ,  $\delta(e_i)$  [72,75-77].

On the other hand, by using the edge (bond)-adjacency relationships we can find other new relation for a molecular graph that will be introduced here. The  $k$ th stochastic edge-adjacency matrix,  $ES^k$  can be obtained directly from  $E^k$ . Here,  $ES^k = [{}^k es_{ij}]$  is a square table of order  $m$  ( $m$  = number of bonds) and the elements  ${}^k es_{ij}$  are defined as follows:

$${}^k es_{ij} = \frac{{}^k e_{ij}}{{}^k SUM(E^k)_i} = \frac{{}^k e_{ij}}{{}^k \delta(e)_i} \quad (1)$$

where,  ${}^k e_{ij}$  are the elements of the  $k$ th power of  $E$  and the SUM of the  $i$ th row of  $E^k$  are named the  $k$ -order edge degree of bond  $i$ ,  ${}^k \delta(e_i)$ . Note that the matrix  $ES^k$  in Eq. 1 has the property that *the sum of the elements in each row* is 1. Such an  $m \times m$  matrix with nonnegative entries having this property is called a “stochastic matrix” [78].

**2.1.1.2. Chemical Information and Bond-based Molecular Vector.** The atom-based molecular vector ( $\bar{x}$ ) used to represent small-to-medium size organic chemicals has been explained in some detail elsewhere [49-68]. In a parallel manner to the development of  $\bar{x}$ , we present the extension to the bond-based molecular vector ( $\bar{w}$ ). The components ( $w_i$ ) of  $\bar{w}$  are numeric values, which represent a certain standard bond property (bond-label). That is to say, these weights correspond to different bond properties for organic molecules. Thus, a molecule having 5, 10, 15, ...,  $m$  bonds can be represented by means of vectors, with 5, 10, 15, ...,  $m$  components, belonging to the spaces  $\mathfrak{R}^5$ ,  $\mathfrak{R}^{10}$ ,  $\mathfrak{R}^{15}$ , ...,  $\mathfrak{R}^m$ , respectively; where  $m$  is the dimension of the real sets ( $\mathfrak{R}^m$ ). This approach allows us

encoding organic molecules such as 2-hydroxybut-2-enitrile through the molecular vector  $\bar{w} = [w_{\text{Csp3-Csp2}}, w_{\text{Csp2=Csp2}}, w_{\text{Csp2-Osp3}}, w_{\text{H-Osp3}}, w_{\text{Csp2-Csp}}, w_{\text{Csp}\equiv\text{Nsp}}]$ . This vector belongs to the product space  $\mathfrak{R}^6$ .

These properties characterize each kind of bond (and bond-types) within the molecule. Diverse kinds of bond weights ( $w_i$ ) can be used in order to codify information related to each bond in the molecule. These bond labels are chemically meaningful numbers such as standard bond distance [79-82], standard bond dipole [79-82] or even mathematical expressions involving atomic weights such as atomic Log P [83], surface contributions of polar atoms [84], atomic molar refractivity [85], atomic hybrid polarizabilities [86], and Gasteiger-Marsilli atomic charge [87], atomic electronegativity in Pauling scale [88] and so on. Here, we characterized each bond with the following parameter:

$$w_i = x_i/\delta_i + x_j/\delta_j \quad (2)$$

which characterizes each bond. In this expression  $x_i$  can be any standard weight of the atom  $i$  bonded with atom  $j$ .  $\delta_i$  is the vertex (atom) degree of atom  $i$ . The use of each scale (bond property) defines alternative molecular vectors,  $\bar{w}$ .

**2.1.1.3. Theory of Non-Stochastic and Stochastic Total (Whole) and Local (Bond and Bond-type) Quadratic Indices.** If a molecule consists of  $m$  bonds (*vector of  $\mathfrak{R}^m$* ), then the  $k$ th total quadratic indices are calculated as quadratic maps (quadratic form) in  $\mathfrak{R}^m$  in canonical basis set. Specifically, the  $k$ th total non-stochastic and stochastic bond-quadratic indices,  $q_k(\bar{w})$  and  ${}^s q_k(\bar{w})$ , are computed from these  $k$ th non-stochastic and stochastic edge adjacency matrices,  $E^k$  and  $ES^k$ , as shown in Eqs. 3 and 4, correspondingly:

$$q_k(\bar{w}) = \sum_{i=1}^m \sum_{j=1}^m {}^k e_{ij} w^i w^j = [\bar{w}]' E^k [\bar{w}] \quad (3)$$

$${}^s q_k(\bar{w}) = \sum_{i=1}^m \sum_{j=1}^m {}^k e_{ij} w^i w^j = [\bar{w}]' ES^k [\bar{w}] \quad (4)$$

where,  $m$  is the number of bonds of the molecule, and  $w^1, \dots, w^m$  are the coordinates of the bond-based molecular vector ( $\bar{w}$ ) in the so-called canonical ('natural') basis. In this basis system, the coordinates of any vector  $\bar{w}$  coincide with the components of this vector [78,89,90]. For that reason, those coordinates can be considered as weights (bond-labels) of the edge of the molecular graph. The coefficients  ${}^k e_{ij}$  and  ${}^k e_{sij}$  are the elements of the  $k$ th power of the matrix  $E(G)$  and  $ES(G)$ , correspondingly, of the molecular graph. The defining equations (3) and (4) for  $q_k(\bar{w})$  and  ${}^s q_k(\bar{w})$ , respectively, can be also written as the single matrix equations (see Eqs. 3 and 4), where  $[\bar{w}]$  is a column vector (an  $m \times 1$  matrix) of the coordinates of  $\bar{w}$  in the canonical basis of  $\mathfrak{R}^m$  and  $[\bar{w}]^t$  (an  $1 \times m$  matrix) is the transpose of  $[\bar{w}]$ . Here,  $E^k$  and  $ES^k$  denote the matrices of quadratic maps with respect to the natural basis set.

In addition to total bond-based quadratic indices, computed for the whole molecule, a local-fragment (bond and bond-type) formalism can be developed. These descriptors are termed local non-stochastic and stochastic quadratic indices,  $q_{kL}(\bar{w})$  and  ${}^s q_{kL}(\bar{w})$ , respectively. The definition of these descriptors is as follows:

$$q_{kL}(\bar{w}) = \sum_{i=1}^m \sum_{j=1}^m {}^k e_{ijL} w^i w^j = [\bar{w}]' E^k_L [\bar{w}] \quad (5)$$

$${}^s q_{kL}(\bar{w}) = \sum_{i=1}^m \sum_{j=1}^m {}^k e_{sijL} w^i w^j = [\bar{w}]' ES^k_L [\bar{w}] \quad (6)$$

where,  $m$  is the number of bonds and  ${}^k e_{ijL}$  ( ${}^k es_{ijL}$ ) is the  $k$ th element of the row “ $i$ ” and column “ $j$ ” of the local matrix  $E_L^k$  ( $ES_L^k$ ). This local matrix is extracted from the  $E^k$  ( $ES^k$ ) matrix and contains information referred to the edges (bonds) of the specific molecular fragments and also of the molecular environment in  $k$  steps. The matrix  $E_L^k$  ( $ES_L^k$ ) with elements  ${}^k e_{ijL}$  ( ${}^k es_{ijL}$ ) is defined as follows:

$$\begin{aligned}
 {}^k e_{ijL} ({}^k es_{ijL}) &= {}^k e_{ij} ({}^k es_{ijL}) \text{ if both } e_i \text{ and } e_j \text{ are edges (bonds) contained within the} \\
 &\quad \text{molecular fragment} \\
 &= \frac{1}{2} {}^k e_{ij} ({}^k es_{ijL}) \text{ if } e_i \text{ and } e_j \text{ are edges (bonds) contained within the molecular} \\
 &\quad \text{fragment but not both} \\
 &= 0 \text{ otherwise}
 \end{aligned} \tag{7}$$

Notice that the above scheme follows the spirit of a Mulliken population analysis [91]. Note also that for every partitioning of a molecule into  $Z$  molecular fragments there will be  $Z$  local molecular fragment matrices. In this case, if a molecule is partitioned into  $Z$  molecular fragments, the matrices  $E^k$  ( $ES^k$ ) can be partitioned into  $Z$  local matrices  $E_L^k$  ( $ES_L^k$ ),  $L = 1, \dots, Z$ , and the  $k$ th power of matrix  $E$  ( $ES$ ) is exactly the sum of the  $k$ th power of the local  $Z$  matrices. In this way, the total non-stochastic and stochastic bond-based quadratic indices are the sum of the non-stochastic and stochastic bond-based quadratic indices, respectively, of the  $Z$  molecular fragments:

$$q_k(\bar{w}) = \sum_{L=1}^Z q_{kL}(\bar{w}) \tag{8}$$

$${}^s q_k(\bar{w}) = \sum_{L=1}^Z {}^s q_{kL}(\bar{w}) \tag{9}$$

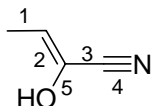
Bond and bond-type quadratic fingerprints are specific cases of local bond-based quadratic indices. In this sense, the  $k$ th bond-type quadratic indices are calculated by adding the  $k$ th bond quadratic indices for all bonds of the same type in the molecule. That is to say, this extension of the bond quadratic index is similar to group additive schemes, in which an index appears for each bond type in the molecule together with its contribution based on the bond quadratic index.

In the bond-type quadratic indices formalism, each bond in the molecule is classified into a bond-type (fragment). In this sense, bonds may be classified into bond types in terms of the characteristics of the two atoms which define the bond. For all data sets, including those with a common molecular scaffold as well as those with very diverse structure, the  $k$ th fragment (bond-type) quadratic indices provide much useful information. Thus, the development of the bond-type quadratic indices description provides the basis for application to a wider range of biological problems in which the local formalism is applicable without the need for superposition or a closely related set of structures. The bond-type descriptors combine three important aspects of structure information: 1) electron accessibility for the bonds of the same type, 2) presence/absence of the bond type, and 3) count of the bonds in the bond type.

It is useful to perform a calculation on a molecule to illustrate the steps in the procedure.

For this, in the next section we depict a pictorial representation of the calculus of the non-stochastic and stochastic quadratic indices of the bond matrix (both total and local) using a simple chemical example.

**2.1.1.4. Sample Calculation.** The quadratic indices of the bond matrix are calculated in the following way. Considering the molecule of 2-hydroxybut-2-enitrile as a simple example, we have the following labelled molecular graph and bond-based adjacency matrices (E and ES). The second ( $k = 2$ ) and third ( $k = 3$ ) power of these matrices and bond-based molecular vector,  $\bar{w}$ , are also given:



$$E^0 = ES^0 = \begin{bmatrix} 1 & & & & \\ & 1 & & & \\ & & 1 & & \\ & & & 1 & \\ & & & & 1 \end{bmatrix} \quad E^1 = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \end{bmatrix} \quad E^2 = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 \\ 0 & 3 & 1 & 1 & 1 \\ 1 & 1 & 3 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 2 \end{bmatrix} \quad E^3 = \begin{bmatrix} 0 & 3 & 1 & 1 & 1 \\ 3 & 2 & 5 & 1 & 4 \\ 1 & 5 & 2 & 3 & 4 \\ 1 & 1 & 3 & 0 & 1 \\ 1 & 4 & 4 & 1 & 2 \end{bmatrix}$$

$$ES^1 = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0.33 & 0 & 0.33 & 0 & 0.33 \\ 0 & 0.33 & 0 & 0.33 & 0.33 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0.5 & 0.5 & 0 & 0 \end{bmatrix} \quad ES^2 = \begin{bmatrix} 0.33 & 0 & 0.33 & 0 & 0.33 \\ 0 & 0.5 & 0.16 & 0.16 & 0.16 \\ 0.16 & 0.16 & 0.5 & 0 & 0.16 \\ 0 & 0.33 & 0 & 0.33 & 0.33 \\ 0.16 & 0.16 & 0.16 & 0.16 & 0.33 \end{bmatrix} \quad ES^3 = \begin{bmatrix} 0 & 0.5 & 0.16 & 0.16 & 0.16 \\ 0.2 & 0.13 & 0.33 & 0.06 & 0.26 \\ 0.06 & 0.33 & 0.13 & 0.2 & 0.26 \\ 0.16 & 0.16 & 0.5 & 0 & 0.16 \\ 0.083 & 0.33 & 0.33 & 0.083 & 0.16 \end{bmatrix}$$

The molecule contains five localized bonds (corresponding to five edges in the H-suppressed molecular graph). To these we will associate the five “bond orbitals” (bond-labels)  $w_1, w_2, w_3, w_4,$  and  $w_5$ . Thus,  $\bar{w} = [w_1, w_2, w_3, w_4, w_5] = [w_{(C-C)}, w_{(C=C)}, w_{(C-C)}, w_{(C=N)}, w_{(C-O)}]$  and each “bond orbitals” can be computed by Eq. 2 using, for instance, the atomic electronegativity in Pauling scale ( $x$ ) [88] as atomic weight (atom-label):

$$w_1 = x_C/1 + x_C/3 = 2.55/1 + 2.55/3 = 3.4$$

$$w_2 = x_C/3 + x_C/4 = 2.55/3 + 2.55/4 = 1.4875$$

$$w_3 = x_C/4 + x_C/4 = 2.55/4 + 2.55/4 = 1.275$$

$$w_4 = x_C/4 + x_N/3 = 2.55/4 + 3.04/3 = 1.650833$$

$$w_5 = x_C/4 + x_O/1 = 2.55/4 + 3.44/1 = 4.0775$$

and therefore,  $\bar{w} = [3.4, 1.4875, 1.275, 1.650833, 4.0775]$ .

Each non-stochastic and stochastic total quadratic indices will have the form:

$$\begin{aligned} \mathbf{q}_k(\bar{w}) = & {}^k e_{11}(w^1)^2 + {}^k e_{21}w^1w^2 + {}^k e_{31}w^1w^3 + {}^k e_{41}w^1w^4 + {}^k e_{51}w^1w^5 + {}^k e_{12}(w^2)^2 + \\ & + {}^k e_{32}w^2w^3 + {}^k e_{42}w^2w^4 + {}^k e_{52}w^2w^5 + {}^k e_{13}w^1w^3 + {}^k e_{23}w^2w^3 + {}^k e_{33}(w^3)^2 + {}^k e_{43}w^3w^4 + \\ & + {}^k e_{53}w^3w^5 + {}^k e_{14}w^1w^4 + {}^k e_{24}w^2w^4 + {}^k e_{34}w^3w^4 + {}^k e_{44}(w^4)^2 + {}^k e_{54}w^4w^5 + {}^k e_{15}w^1w^5 + \\ & + {}^k e_{25}w^2w^5 + {}^k e_{35}w^3w^5 + {}^k e_{45}w^4w^5 + {}^k e_{55}(w^5)^2 = \sum_{(i)} {}^k e_{ii} (w^i)^2 + 2 \sum_{(i,j)} {}^k e_{ij} w^i w^j \quad (10) \end{aligned}$$

$$\begin{aligned} {}^s \mathbf{q}_k(\bar{w}) = & {}^k es_{11}(w^1)^2 + {}^k es_{21}w^1w^2 + {}^k es_{31}w^1w^3 + {}^k es_{41}w^1w^4 + {}^k es_{51}w^1w^5 + {}^k es_{12}w^1w^2 + \\ & + {}^k es_{22}(w^2)^2 + {}^k es_{32}w^2w^3 + {}^k es_{42}w^2w^4 + {}^k es_{52}w^2w^5 + {}^k es_{13}w^1w^3 + {}^k es_{23}w^2w^3 + \\ & + {}^k es_{33}(w^3)^2 + {}^k es_{43}w^3w^4 + {}^k es_{53}w^3w^5 + {}^k es_{14}w^1w^4 + {}^k es_{24}w^2w^4 + {}^k es_{34}w^3w^4 + \\ & + {}^k es_{44}(w^4)^2 + {}^k es_{54}w^4w^5 + {}^k es_{15}w^1w^5 + {}^k es_{25}w^2w^5 + {}^k es_{35}w^3w^5 + {}^k es_{45}w^4w^5 + \\ & + {}^k es_{55}(w^5)^2 = \sum_{(i)} {}^k es_{ii} (w^i)^2 + 2 \sum_{(i,j)} {}^k es_{ij} w^i w^j \quad (11) \end{aligned}$$

The  ${}^k e_{ii}$ 's and  ${}^k es_{ii}$ 's can be considered a measure of the attraction of an electron for a bond in the  $k$  step. The  ${}^k e_{ij}$ 's and  ${}^k es_{ij}$ 's are the terms of interaction between two bonds in the  $k$  step. The  ${}^k e_{ij}$ 's =  ${}^k e_{ji}$ 's are equal by symmetry (non-oriented molecular graph), however,  ${}^k es_{ij} \neq {}^k es_{ji}$ . This is a logical result because the  $k$ th  $es_{ij}$  elements are the transition probabilities with the 'electrons' moving from bond  $i$  to  $j$  at the discrete time periods  $t_k$  and it should be different in both senses. This result is in total agreement if the electronegativity of the two atom types in the bonds are taken into account.

In this way,  $E^k$  and  $ES^k$  can be seen as graph-theoretic electronic-structure models [92]. In fact, quantum chemistry starts from the fact a molecule is made up of electrons

and nuclei. The distinction here between bonded and non-bonded atoms is difficult to justify. Any two nuclei of a molecule interact directly and indirectly through the electrons present in the molecule. Only the intensity of this interaction varies in going from one pair of nuclei to another. In this sense, the electron in an arbitrary bond  $i$  can move (step-by-step) to other bonds at different discrete time periods  $t_k$  ( $k = 0, 1, 2, 3, \dots$ ) through the chemical-bonding network. That is to say, the  $E^1$  and  $ES^1$  matrices consider the valence-bond electrons in one step and their power ( $k = 0, 1, 2, 3, \dots$ ) can be considering as an interacting–electron chemical–network model in  $k$  step. This model can be seen as an intermediate between the quantitative quantum-mechanical Schrödinger equation and classical chemical bonding ideas [92].

On the other hand, the  $k$ th ( $k = 0 - 3$ ) non-stochastic total quadratic indices can be expressed as the sum of the local (bond) quadratic indices for this molecule as follows:

$$q_0(\bar{w}) = q_{0L}(\bar{w})_1 + q_{0L}(\bar{w})_2 + q_{0L}(\bar{w})_3 + q_{0L}(\bar{w})_4 + q_{0L}(\bar{w})_5 = 11.56 + 2.21265625 + 1.625625 + 2.72525069 + 16.6260063 = 34.7495382$$

$$q_1(\bar{w}) = q_{1L}(\bar{w})_1 + q_{1L}(\bar{w})_2 + q_{1L}(\bar{w})_3 + q_{1L}(\bar{w})_4 + q_{1L}(\bar{w})_5 = 5.0575 + 13.0193438 + 9.2001875 + 2.1048125 + 11.2640938 = 40.6459375$$

$$q_2(\bar{w}) = q_{2L}(\bar{w})_1 + q_{2L}(\bar{w})_2 + q_{2L}(\bar{w})_3 + q_{2L}(\bar{w})_4 + q_{2L}(\bar{w})_5 = 29.7585 + 17.0554271 + 16.30725 + 11.9121382 + 65.1108792 = 140.144194$$

$$q_3(\bar{w}) = q_{3L}(\bar{w})_1 + q_{3L}(\bar{w})_2 + q_{3L}(\bar{w})_3 + q_{3L}(\bar{w})_4 + q_{3L}(\bar{w})_5 = 38.9838333 + 55.7973646 + 44.17875 + 21.1141583 + 98.9031604 = 258.977267$$

The terms in the summations for calculating the total quadratic indices are the so-called local (bond) quadratic indices. We have written these terms in the consecutive order of the bond labels in the graph. For instance, the non-stochastic bond quadratic



indices of order 0, 1, 2 and 3 for the bond labelled as 1 are 11.56, 5.0575, 29.7585, and 38.9838333, respectively.

The  $k$ th total bond-based stochastic quadratic indices values are also the sum of the  $k$ th local (bond) stochastic quadratic indices values for all bonds in the molecule:

$${}^s q_0(\bar{w}) = {}^s q_{0L}(\bar{w})_1 + {}^s q_{0L}(\bar{w})_2 + {}^s q_{0L}(\bar{w})_3 + {}^s q_{0L}(\bar{w})_4 + {}^s q_{0L}(\bar{w})_5 = 11.56 + 2.21265625 + 1.625625 + 2.72525069 + 16.6260063 = 34.7495382$$

$${}^s q_1(\bar{w}) = {}^s q_{1L}(\bar{w})_1 + {}^s q_{1L}(\bar{w})_2 + {}^s q_{1L}(\bar{w})_3 + {}^s q_{1L}(\bar{w})_4 + {}^s q_{1L}(\bar{w})_5 = 3.37166667 + 6.53105469 + 4.20156771 + 1.40320833 + 4.6933724 = 20,2008698$$

$${}^s q_2(\bar{w}) = {}^s q_{2L}(\bar{w})_1 + {}^s q_{2L}(\bar{w})_2 + {}^s q_{2L}(\bar{w})_3 + {}^s q_{2L}(\bar{w})_4 + {}^s q_{2L}(\bar{w})_5 = 8.40295833 + 3.04720573 + 3.079125 + 3.20513877 + 12.5680443 = 30.3024721$$

$${}^s q_3(\bar{w}) = {}^s q_{3L}(\bar{w})_1 + {}^s q_{3L}(\bar{w})_2 + {}^s q_{3L}(\bar{w})_3 + {}^s q_{3L}(\bar{w})_4 + {}^s q_{3L}(\bar{w})_5 = 4.94428472 + 4.80340608 + 3.65101563 + 2.80005408 + 8.72457578 = 24.9233363$$

## 2.2. Computational Strategies.

The main steps for the application of present method in QSAR/QSPR and drug design can be briefly summarized in the following set of steps: 1) Draw the molecular pseudographs for each molecule of the data set, using the software drawing mode. This procedure is performed by a selection of the active atomic symbol belonging to the different groups in the periodic table of the elements, 2) Use appropriated atomic properties in order to weight and differentiate the molecular bonds. In this study, the weights used are those previously proposed for the calculation of the DRAGON descriptors, [88, 94-95] i.e., atomic mass (M), atomic polarizability (P), atomic Mulliken electronegativity (K), van der Waals atomic volume (V), plus the atomic electronegativity in Pauling scale (G) [96]. The values of these atomic labels are shown

in Table 1. [88,94-95]. In this case, we are used the mathematical expression in Eq. 2, which involving atomic weights (see Table 1), 3) Compute the total and local (bond and bond-type) non-stochastic and stochastic quadratic indices. It can be carried out in the software calculation mode, where you can select the atomic properties and the descriptor family previously to calculate the molecular indices. This software generates a table in which the rows correspond to the compounds, and columns correspond to the total and local bond-based quadratic indices or other molecular descriptors family implemented in this program, 4) Find a QSPR/QSAR equation by using several multivariate analytical techniques, such as multilinear regression analysis (MRA), neural networks (NN), linear discrimination analysis (LDA), and so on. That is to say, we can find a quantitative relation between an activity **A** and the linear indices having, for instance, the following appearance,  $\mathbf{A} = a_0q_0(\bar{w}) + a_1q_1(\bar{w}) + a_2q_2(\bar{w}) + \dots + a_kq_k(\bar{w}) + c$ , where **A** is the measured activity,  $q_k(\bar{w})$  are the *k*th total bond-based quadratic indices, and the  $a_k$ 's are the coefficients obtained by the linear regression analysis, 5) Test the robustness and predictive power of the QSPR/QSAR equation by using internal (cross-validation) and external (using a test set and an external predicting set) validation techniques, and 6) Apply the obtained LDA-based QSAR models as cheminformatic tool for identifying leads through ligand-based virtual screening-drug discovery process.

The bond-based **TOMOCOMD-CARDD** descriptors computed in this study were the following:

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

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

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

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

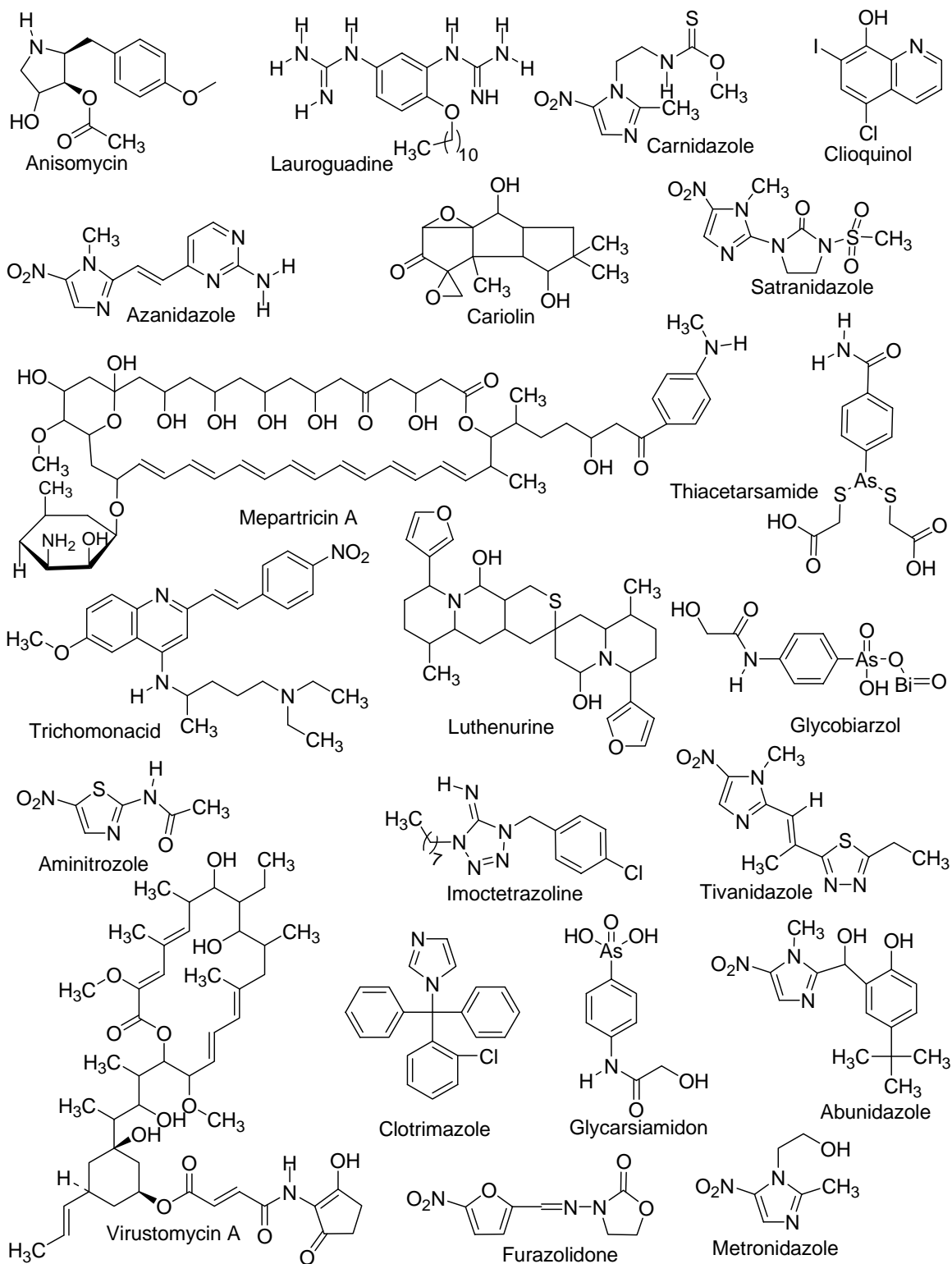
**Table 1.** Values of the Atom Weights Used for Quadratic Indices Calculation [88,94-95].

ID	Atomic Mass	VdW Volume	Mulliken Electronegativity	Polarizability	Pauling Electronegativity
H	1.01	6.709	2.592	0.667	2.2
B	10.81	17.875	2.275	3.030	2.04
C	12.01	22.449	2.746	1.760	2.55
N	14.01	15.599	3.194	1.100	3.04
O	16.00	11.494	3.654	0.802	3.44
F	19.00	9.203	4.000	0.557	3.98
Al	26.98	36.511	1.714	6.800	1.61
Si	28.09	31.976	2.138	5.380	1.9
P	30.97	26.522	2.515	3.630	2.19
S	32.07	24.429	2.957	2.900	2.58
Cl	35.45	23.228	3.475	2.180	3.16
Fe	55.85	41.052	2.000	8.400	1.83
Co	58.93	35.041	2.000	7.500	1.88
Ni	58.69	17.157	2.000	6.800	1.91
Cu	63.55	11.494	2.033	6.100	1.9
Zn	65.39	38.351	2.223	7.100	1.65
Br	79.90	31.059	3.219	3.050	2.96
Sn	118.71	45.830	2.298	7.700	1.96
I	126.90	38.792	2.778	5.350	2.66

### 2.3. Database Selection

In order to obtain mathematical expressions capable of discriminating between active and inactive compounds, the chemical information contained in a great number of compounds with and without the desired biological activity must be statistically processed. Taking into account that the most critical aspect in the construction of a training data set is the molecular diversity of the included compounds, we selected a group of 123 organic chemicals having as much structural variability as possible. The 50 antitrichomonals considered in this study are representative of families with diverse structural patterns and action modes. Figure 1 shows a representative sample of such active compounds. On the other hand, 73 compounds having different clinical uses were selected for the set of inactive compounds, through a random selection, guaranteeing also a great structural variability. All these chemicals were taken from the Negwer Handbook [97], and Merck Index [98], where their names, synonyms and structural formulas can be found. From these 123 chemicals, 91 were chosen at random to form the training set, being 40 of them active and 51 inactive ones.

The great structural variability of the selected training data set makes possible the discovery of lead compounds, not only with determined mechanisms of antitrichomonal activity, but also with novel modes of action (which will be illustrated well in this paper in a virtual experiment for lead compounds generation). The remaining subseries consisting of 10 trichomonacidal and 22 non-trichomonacidal were prepared as test sets for the external validation of the models (32 chemicals). These compounds were never used in the development of the classification models.



**Figure 1.** Random sample of the molecular families of trichomonacidal agents studied here.

#### 2.4. Data Analysis and Processing: *Linear Discriminant Analysis*.

The discriminant functions were obtained by using the Linear Discriminant Analysis (LDA) [99] as implemented in the STATISTICA [100]. The default parameters of this program were used in the development of the model. Forward stepwise was fixed as the strategy for variable selection. The principle of parsimony (Occam's razor) was taken into account as they strategy for model selection. In its original form, the Occam's razor states that »*Numquam ponenda est pluritas sin necessitate*«, which can be translated as »Entities should not be multiplied beyond necessity« [101]. In this case, simplicity is loosely equated with the number of parameters in the model. If we understand the predictive error to be the error rate for unseen examples, the Occam's razor can be stated for the selection of QSAR/QSPR models as ("*QSAR/QSPR Occam's Razor*") : Given two QSAR/QSPR models with the same predictive error, the simplest one should be preferred because simplicity is desirable in itself [101]. In this connection, we select the model with higher statistical signification but having as few parameters ( $a_k$ ) as possible.

The quality of the models were determined by examining Wilks'  $\lambda$  parameter ( $U$ -statistic), squared Mahalanobis distance ( $D^2$ ), Fisher ratio (F) and the corresponding  $p$ -level ( $p(F)$ ) as well as the percentage of good classification in the training and test sets [99]. Models with a proportion between the number of cases and variables in the equation lower than 5 were rejected.

The Wilks'  $\lambda$  for the overall discrimination can take values in the range of 0 (perfect discrimination) to 1 (no discrimination). The  $D^2$  statistics indicates the separation of the respective groups, showing whether the model possesses an appropriate discriminatory power for differentiating between the two respective groups.

By using the models, one compound can then be classified as either active, if  $\Delta P\% > 0$ , being  $\Delta P\% = [P(\text{Active}) - P(\text{Inactive})] \times 100$  or inactive otherwise.  $P(\text{Active})$  and  $P(\text{Inactive})$  are the probabilities with which the equations classify a compound as active and inactive, respectively.

The statistical robustness and predictive power of the obtained model were assessed using a prediction (test) set [102]. Also a leave-group-out (LGO) cross-validation strategy was carried out. In this case, 10% of the data set was used as group size, i.e. groups including 10% of the training data set were left out and predicted by the model based on the remaining 90%. This process was carried out 10 times on 10 unique subsets. In this way, every observation was predicted once (in its group of left-out observations). The overall mean for this process (10% *full* leave-out cross-validation) was used as a good indication of robustness, stability and predictive powers of the obtained models [102].

Finally, the calculation of percentages of global good classification (accuracy), sensibility, specificity (also known as ‘hit rate’), false positive rate (also known as ‘false alarm rate’) and Matthews correlation coefficient ( $C$ ) in the training and test sets permitted the assessment of the model [103].

## **2.5. Determination of *in vitro* Trichomonacidal Activity**

The biological activity was assayed on Tv JH31A #4 Ref. No. 30326 (ATCC, MD, USA) in modified Diamond medium supplemented with equine serum and grown at 37 °C (5% CO<sub>2</sub>). 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. MTZ (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 cytotoxic (%C) and cytostatic activities (%CA), in relation to controls as previously reported [104.105].

## 2.6. Determination of *in vivo* Trichomonacidal Activity

In this study, we examined the protective efficacy of G-1 in ovariectomized rats and dealt with estradiol [106]. The used excipient to dissolve the active principle G-1 was Migliol® 810 N (HULS AG Canada).

The taking of sample for the diagnosis 1 of the infection were carried out previous to the application intravaginal of the first dose with G-1, this allowed us to know the number of rats infected before the first treatment. The days 7 and 8 were carried out a similar procedure that it allowed us to value the effect of the first one and second dose (see Table 2)

**Table 2.** Infection with *T. vaginalis*, treatment with G-1 and diagnosis of the rats.

Group	Day								
	0	2 - 3	6		7		8		9; 11; 13; 15
	Estrogen.	Infection	Diag	T. 1	Diag	T. 2	Diag	T. 3	Diag
I				G-1		G-1		G-1	
II	Estradiol 10 mg/kg	<i>T.v</i>	1	M	2	M	3	M	4 - 7
III				-		-		-	
IV				Mtz		Mtz		Mtz	

T. = Treatment number

Estrogen. = Estrogenization

Diag = Diagnostic number

*T.v.* =  $5 \times 10^6$  *T.v*/ml /intravaginal inoculation 200  $\mu$ l

M = Migliol

G-1 = 2-bromo-5-(2-bromo-2-nitrovinil)furano /0.125%



### 3. RESULTS AND DISCUSSION

#### 3.1. Development and Validation of the Discriminant Functions.

Although the number of existing statistical methods to get classification functions is relatively extensive, we select linear discriminant analysis (LDA) given the simplicity of the method [99]. The use of LDA in rational drug design has been extensively reported by different authors [46,52-54,56-64]. Therefore, LDA was also the technique used in the generation of discriminant functions in the current work. Making use of the LDA technique implemented in the STATISTICA software [100], the following linear models were obtained; in which total as well as local non-stochastic and stochastic bond-based quadratic indices were used as independent variables:

$$\begin{aligned} \text{Class} = & -4.92 - 1.40 \times 10^{-3} q_1^M(\bar{w}) + 0.22 q_0^H(\bar{w}) + 2.01 \times 10^{-2} q_1^P(\bar{w}) - 1.44 \times 10^3 q_{0L}^V(\bar{w}_E) \\ & + 0.13 q_{0L}^E(\bar{w}_E) + 5.82 \times 10^{-2} q_{1L}^H(\bar{w}_E) \end{aligned} \quad (12)$$

$$N = 91 \quad \lambda = 0.44 \quad D^2 = 5.02 \quad F(6.84) = 17.71 \quad p < 0.0001$$

$$\begin{aligned} \text{Class} = & -5.50 + 2.78 \times 10^{-3} q_{0L}^{Ms}(\bar{w}_E) + 7.87 \times 10^{-3} q_{1L}^{Ms}(\bar{w}_E) - 1.49 \times 10^{-2} q_{3L}^{Ms}(\bar{w}_E) \\ & - 0.25 q_{1L}^{Ps}(\bar{w}_E) + 0.49 q_{3L}^{Es}(\bar{w}_E) \end{aligned} \quad (13)$$

$$N = 91 \quad \lambda = 0.36 \quad D^2 = 6.86 \quad F(5.85) = 29.39 \quad p < 0.0001$$

where N is the number of compounds,  $\lambda$  is Wilks' statistics,  $D^2$  is the square of the Mahalanobis distance, F is the Fisher ratio and  $p$  is the significance level.

Model **12** classifies correctly 85% of active and 90.20% of inactive compounds in the training set for a global good classification (accuracy) of 87.91%. Model **13** classifies correctly 89.01% of the compounds in training set. Specifically, the model correctly classifies 35 out of 40 (87.50%) trichomonacidal compounds and 46 out of 51 (90.20%)

inactive chemicals in the training series. On the other hand, Eqs. **12** and **13** show a 87.50% (30/32) and 84.38% (27/32) of global predictability in the prediction series, respectively. These results validate the models for use in the ligand-based virtual screening taking into consideration that 85.0% is considered as an acceptable threshold limit for this kind of analysis [107].

In Tables 3 and 4 we give the names of all compounds in the training and test active and inactive sets together with their posterior probabilities calculated from the Mahalanobis distance using both equations. The same information of all compounds in the training and test inactive set appears in Table 5 which summarizes the results of the classifications for both models in the training and test groups.

A more serious analysis was carried out by calculating most of the parameters commonly used in medical statistics (accuracy, sensitivity, specificity and false positive rate) and the Matthews correlation coefficient (*C*). Table 5 also lists these parameters for both obtained models [103,108]. While the sensitivity is the probability of correctly predicting a positive example, the specificity is the probability that a positive prediction is correct. On the other hand, *C* quantifies the strength of the linear relation between the molecular descriptors and the classifications, and it may often provide a much more balanced evaluation of the prediction than, for instance, the percentages [103,108]. The obtained models, Eqs. **12** and **13**, showed a high *C* of 0.75 (0.71) and 0.78 (0.65) in training (test) sets, correspondingly.

Although, the most important criterion for the quality of the discriminant model is based on the statistics for the external prediction set, for a more exhaustive testing of the

predictive power of the models, we carried out a leave-10-*fold* full-out (LGO) cross-validation procedure. For each group of observations left out (10% of the whole data set, 9 compounds), a model was developed from the remaining 90% of the data (81 compounds). This process was carried out ten times on ten unique subsets. The statistical results are depicted in Table 6. The overall mean of the correct classification in training (test) set for this process for Eq. **12** and **13** was 87.69% (85.64%) and 89.03% (87.86%), correspondingly. The result of predictions on the 10% *full* cross-validation test evidenced the quality (robustness, stability and predictive power) of the obtained models.

**Table 3.** Names and classification of active compounds in training and test series according to the two *TOMOCOMD-CARDD* models developed in this work.

Name	$\Delta P\%$ <sup>a</sup>	$\Delta P\%$ <sup>b</sup>	Name	$\Delta P\%$ <sup>a</sup>	$\Delta P\%$ <sup>b</sup>
Active training set					
Anisomycin	<b>-29.28</b>	38.15	Abunidazole	32.45	93.47
Virustomycin A	78.19	99.40	Imoctetrazoline	35.27	<b>-36.35</b>
Azanidazole	92.53	98.84	Forminitrazole	89.57	83.54
Carnidazole	94.20	92.49	Chlomizol	90.67	98.19
Propenidazole	97.42	99.67	Acinitrazole	89.22	75.81
Lauroguadine	<b>-92.30</b>	<b>-45.38</b>	Moxnidazole	99.93	99.99
Mepartricin A	91.53	98.36	Isometronidazole	68.48	91.48
Metronidazole	69.08	91.09	Mertronidazole phosphate	90.49	71.43
Nifuratel	99.22	99.97	Benzoylmetronidazole	98.17	98.76
Nifuroxime	89.84	98.08	Bamnidazole	96.73	96.93
Nimorazole	90.55	93.45	Glycarsiamidon	<b>-34.15</b>	<b>-10.61</b>
Secnidazole	59.89	77.39	Fexinidazole	81.25	99.62
Cariolin	<b>-84.08</b>	6.12	Piperanitrozole	96.83	84.34
2 -Amino -5 -nitrothiazola	48.25	39.79	Gynotabs	87.94	84.24
Glycobiartzol	83.73	97.80	Pirinidazole	90.84	96.71
Clioquinol	37.59	98.07	Metronidazole hydrogen succinate	98.96	97.36
Diiodohydroxy quinoline	63.69	95.97	Tolamizol	96.88	93.24
Ornidazol	92.97	98.30	Thiacetarsamide	6.22	<b>-65.88</b>
Trichomonacid	87.66	92.43	Tivanidazole	95.85	87.49
Lutenurine	<b>-31.81</b>	51.26	Policresulen	<b>-63.25</b>	<b>-78.01</b>
Active test set					
Acertarsona	<b>-27.20</b>	10.11	Pentamycin	<b>-96.53</b>	<b>-99.57</b>
Furazolidone	99.72	99.93	Azomycin	56.06	75.54
Mepartricin B	93.76	50.32	Ternidazole	66.21	80.99
Aminitrozole	89.22	75.81	Misonidazole	74.42	94.52
Clotrimazol	1.93	-37.19	Satranidazole	98.48	98.93

<sup>a,b</sup>Antitrichomonal activity predicted by Eqs (12) and (13), respectively:  $\Delta P\% = [P(\text{Active}) - P(\text{Inactive})] \times 100$ .

**Table 4.** Names and classification of inactive compounds in training and test series according to the two *TOMOCOMD-CARDD* models developed in this work.

Name	$\Delta P\%$ <sup>a</sup>	$\Delta P\%$ <sup>b</sup>	Name	$\Delta P\%$ <sup>a</sup>	$\Delta P\%$ <sup>b</sup>
Inactive training set					
Amantadine	-99.81	-99.56	Nonaferone	-71.30	-95.07
Thiacetazone	-78.46	-98.10	Rolipram	-77.60	-56.38
Cloral betaine	-99.28	-99.86	N-hydroxymethyl-N-methylurea	-98.25	-99.87
Carbavin	-89.39	-90.44	4 chlorobenzoic acid	-70.71	-86.21
Norantoin	-66.14	-76.73	Acetanilide	-93.65	-98.37
Orotonsan Fe	-14.22	<b>20.01</b>	Guanazole	-98.65	-99.71
Picosulfate	<b>84.58</b>	<b>68.69</b>	Tetramin	-97.88	-99.37
Naftazone	-67.94	-85.64	Mecysteine	-98.74	-98.30
Besunide	-70.54	-85.55	Cirazoline	-87.74	-94.17
Acetazolamide	-12.85	-91.01	Methocarbamol	-25.86	-11.53
Propamine"soviet	-99.91	-99.99	Lysergide	-87.90	-87.90
RMI 11894	-99.12	-99.43	Dopamine	-98.72	-97.60
Ag 307	-94.12	-99.45	Bufeniode	<b>11.15</b>	-93.94
Barbismethylii iodide	-99.16	-97.44	Celiprolol	-65.72	-86.20
Pancuronium bromide	-96.31	-96.64	Erysimin	-40.42	-18.58
Vinyl ether	-92.23	-97.95	Peruvoside	-13.71	64.09
Basedol	-61.43	-79.61	Amitraz	-89.33	-93.66
Carbimazole	32.19	-31.10	Proclonol	-59.16	-94.61
Didym levulinate	-94.02	-98.80	Asame	-81.15	-97.62
Perchloroethane	-99.51	-97.54	KC-8973	-51.60	-88.11
Pyrantel tartrate	-93.30	-98.23	Ethydine	<b>38.73</b>	<b>11.75</b>
Fentanyl	-71.40	-96.78	Magnesii metioglicas	-76.55	-99.99
Petidina	-92.06	-90.11	Alibendol	-88.74	-50.31
Tenalidine tartrate	-99.27	-99.91	Diponium Bromide	-94.93	-97.04
Bamipine	-96.60	-98.94	Streptomycin	<b>74.73</b>	<b>86.40</b>
Colestipol	-99.82	-99.91			
Inactive test set					
Citenazone	-79.26	-99.27	Metriponate	-96.96	-78.40
Methenamine	-83.69	-90.40	Ciclopramine	-97.13	-92.79
Pentrichloral	-93.97	<b>73.04</b>	Litracen	-99.63	-99.09
Calcium Sodium ferriclate	-100.00	-100.00	Trimetilsulfonium hidroxide	-99.98	-100.00
Ferroceron	90.05	-97.27	Norgamem	-96.22	-97.12
Emodin	-76.24	-79.33	Emylcamate	-91.15	-95.42
Butanolum	-99.09	-99.60	Acetylcholine	-99.40	-99.74
Spirolactone	-86.18	-93.07	Carazolol	-91.64	-96.38
Bromcholine	-99.95	<b>74.90</b>	Cefazolin	<b>99.69</b>	<b>99.69</b>
Imekhin	-99.77	-99.64	Penicillin I	-33.19	-52.65
Diphenadione	-70.80	-94.80	Aziromycin	-89.86	-92.22

<sup>a,b</sup>Antitrichomonal activity predicted by Eqs 12 and 13, respectively:  $\Delta P\% = [P(\text{Active}) - P(\text{Inactive})] \times 100$ .

**Table 5.** Prediction performances for two LDA-based QSAR models (using non-stochastic and stochastic bond-type quadratic indices) in the training and test sets.

	Matthews Corr. Coefficient (C)	Accuracy	'Q <sub>Total</sub> ' (%)	Sensitivity 'hit rate' (%)	Specificity (%)	False positive rate 'false alarm rate' (%)
<b>Non-Stochastic Bond-Type Quadratic Indices (Eq. 12)</b>						
Learning set	0.75	87.91		0.85	0.87	0.098
Predicting set	0.71	87.50		0.8	0.80	0.09
<b>Stochastic Bond-Type Quadratic Indices (Eq. 13)</b>						
Learning set	0.78	89.01		0.88	0.88	0.10
Predicting set	0.65	84.38		0.8	0.73	0.14

**Table 6.** Results of the 10-*fold* full cross-validation procedure.

Groups	Q% <sup>a</sup>	$\lambda$	D <sup>2</sup>	F	Q% <sup>b</sup>	Q% <sup>a</sup>	$\lambda$	D <sup>2</sup>	F	Q% <sup>b</sup>
<b>Eq. 12 (Non-Stochastic Bond-based Quadratic indices)</b>						<b>Eq. 13 (Stochastic Bond-based Quadratic indices)</b>				
1	90.12	0.42	5.52	17.24	80.00	90.12	0.34	7.72	29.32	80.00
2	87.80	0.43	5.15	16.24	88.89	89.02	0.37	6.89	26.44	77.78
3	85.37	0.47	4.54	14.33	100.00	87.80	0.38	6.49	24.92	100.00
4	86.59	0.47	4.43	13.98	100.00	87.80	0.39	6.16	23.63	100.00
5	87.80	0.44	5.03	15.88	77.78	89.02	0.37	6.71	25.74	88.89
6	87.80	0.41	5.74	18.11	88.89	90.24	0.35	7.37	28.28	77.78
7	86.59	0.42	5.47	17.27	77.78	90.24	0.36	7.12	27.33	77.78
8	89.02	0.44	5.09	16.07	77.78	89.02	0.36	6.96	26.72	100.00
9	89.02	0.43	5.32	16.79	77.78	89.02	0.36	7.17	27.50	88.89
10	86.75	0.44	5.06	16.23	87.50	87.95	0.37	6.75	26.33	87.50
Mean	87.69	0.44	5.14	16.21	85.64	89.03	0.36	6.93	26.62	87.86
SD	1.42	0.02	0.41	1.28	8.92	0.96	0.02	0.44	1.64	9.49

<sup>a, b</sup> Global good classification from both models in training (90% of the data) and test (10% of the data) sets, respectively.

### 3.2. 'Virtual' and 'in Silico' Screening as Promissory Alternative for Drug Discovery

In addition to high-throughput screening technology, virtual (*in silico*) screening has become one of the main tools for identifying leads [47,48,109]. Virtual screening is actually one of the computational tools used to filter out unwanted chemicals from physical and/or *in silico* libraries [47,48,109]. Virtual screening techniques may be classified according to their particular modeling of molecular recognition and the type of algorithm used in database searching [47,48,109]. If the target (or at least its active site) 3D structure is known, one of the structure-based virtual screening methods can be

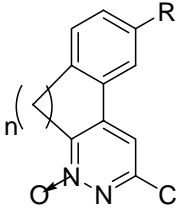
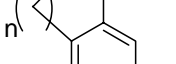
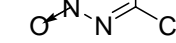
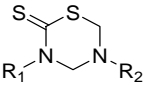
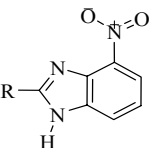
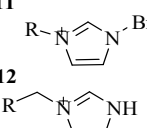
applied. By contrast, ligand-based methods are founded on the principle of similarity, that is, similar compounds are assumed to produce similar effects. The absence of a receptor 3D structure is the main reason for the application of ligand-based methods. However, most (Q)SAR methods are focused on one single family of compounds or a specific action mode. Nevertheless, our group has shown that new lead drugs can be designed and/or selected even if their mechanism of action is completely unknown, by using algorithms based on the structural characterization of a structurally diverse database with molecular descriptors and some pattern recognition technologies such as LDA [46,58,59,110,111].

### **3.2.1. Corroboration of the models' predictive power through a second external test set.**

In order to prove the possibilities of the present approach for the ligand-based virtual screening of antitrichomonal compounds, we have selected a series of 12 compounds, as a second external test set, whose activities against *Tv* have already been proved by several researchers [112,113,114]. They all were evaluated with models **12** and **13** as active/inactive ones. Its structures as well as the results of the classification are shown in Table 7.

By these means, the present study is conducted to test the possibilities of the classification models developed here, in detecting trichomonacidal with diverse chemical structures. The verification of the predictions carried out by all the obtained models comes from the recent reports in the literature, from where these compounds were selected. The results of the classification of the compounds in this external test set are also summarized in Table 7.

**Table 7.** Lead identification among chemicals extracted from literature as active or inactive toward the antitrichomonal activity by using LDA-based QSAR models in simulate virtual screening.

		R <sub>1</sub>	R <sub>2</sub>	
 <p>1: n = CH<sub>2</sub> R = H</p>	 <p>2: n = CH<sub>2</sub>-CH<sub>2</sub> R = H</p>	 <p>3: n = CH<sub>2</sub>=CH<sub>2</sub> R = NH<sub>2</sub></p>	<p><b>4-10</b></p> 	
			 <p><b>11</b></p>	 <p><b>12</b></p>
			<p><b>4</b></p> <p>furfuril</p>	<p>(CH<sub>2</sub>)<sub>5</sub>-COOH</p>
	<p><b>5</b></p> <p>furfuril</p>	<p>CH<sub>2</sub>-COOH</p>		
	<p><b>6</b></p> <p>furfuril</p>	<p>CH(CH<sub>2</sub>Ph)-COOH</p>		
	<p><b>7</b></p> <p>furfuril</p>	<p>CH<sub>2</sub>-CONH-CH<sub>2</sub>-COOH</p>		
	<p><b>8</b></p> <p>furfuril</p>	<p>CH(CH<sub>2</sub>CONH<sub>2</sub>)-COOH</p>		
	<p><b>9</b></p> <p>furfuril</p>	<p>CH[CH<sub>2</sub>-CH(CH<sub>3</sub>)<sub>2</sub>]-COOH</p>		
	<p><b>10</b></p> <p>(CH<sub>2</sub>)<sub>5</sub>-COOH</p>	<p>(CH<sub>2</sub>)<sub>5</sub>-COOH</p>		

Comp. <sup>a</sup>	Ref. <sup>b</sup>	ΔP% <sup>c</sup>	ΔP% <sup>d</sup>	Antitrichomonal activity
1	Gavini et. al., 2000	-14.70	-36.09	inactive
2		-18.99	-42.54	inactive
3		-22.78	-24.75	inactive
4	Ochoa et al., 1999	66.49	<b>-28.97</b>	100 μg/ml = 100 <sup>e</sup> 10 μg/ml = (100) <sup>f</sup> 1 μg/ml = (100) <sup>f</sup>
5		76.66	42.66	100 μg/ml = 100 <sup>e</sup> 10 μg/ml = (100) <sup>f</sup> 1 μg/ml = (97) <sup>f</sup>
6		86.39	79.25	100 μg/ml = 100 <sup>e</sup> 10 μg/ml = (18) <sup>f</sup> 1 μg/ml = (12) <sup>f</sup>
7		91.03	60.63	100 μg/ml = 100 <sup>e</sup> 10 μg/ml = (100) <sup>f</sup> 1 μg/ml = (73) <sup>f</sup>
8		87.67	83.14	100 μg/ml = 100 <sup>e</sup> 10 μg/ml = (100) <sup>f</sup> 1 μg/ml = (93) <sup>f</sup>
9		68.30	72.92	100 μg/ml = 100 <sup>e</sup> 10 μg/ml = (33) <sup>f</sup> 1 μg/ml = (94) <sup>f</sup>
10		<b>-5.25</b>	<b>-97.46</b>	100 μg/ml = 100 <sup>e</sup> 10 μg/ml = (25) <sup>f</sup> 1 μg/ml = (65) <sup>f</sup>
11	Alcalde et. al., 1995	<b>90.34</b>	<b>83.38</b>	Inactive
12		83.68	73.30	MLC <sup>g</sup> = 50 μg/ml LD <sub>50</sub> <sup>h</sup> = 50 μg/ml

<sup>a</sup>The molecular structures of the compounds represented with numbers are shown at the top of this table.

<sup>b</sup>Bibliographical references from where molecules together with its *in vitro* activities were taken.

<sup>c,d</sup>Antitrichomonal activity predicted by Eq 12 and Eq 13; ΔP% = [P(Active) - P(Inactive)]x100.

<sup>e</sup>Percentage of reduction of *T. Vaginalis* or cytotoxic activity at the indicated doses at 24h. <sup>f</sup>Specific activity against *T. Vaginalis* (in brackets) expressed as percentages of growth inhibition or cytostatic activity at 24h.

<sup>g</sup>MLC: minimum lethal concentration that killed all the parasites by 24h. <sup>h</sup>LD<sub>50</sub>: minimum concentration that reduced the number of parasites at least 50%.

As can be seen, both models classify correctly most of the 12 selected compounds. The first model (Eq. **12**) classifies only two lead-compound incorrectly (one of them as false positive and the other as false negative) for yielding 83.33% of correct classification, while the second model (Eq. **13**) classifies three lead-compound incorrectly (two of them as false positive and the other as false negative) for yielding 75.00% of correct classification. This result is the most important validation for the models developed here since it has been able to detect a series of compounds as active from a database composed of compounds selected from literature and these chemicals have shown the predicted activity.

The next step in this approach would be the inclusion of these ‘novel’ compounds in the training set and the developing of a new discrimination model. This new model can be significantly different from the previous one, due to the inclusion of a new structural pattern, but it should be able to recognize a greater number of such compounds as trichomonacidal. By these ways, the derivation of the classifier model is considered as an iterative process, in which novel compounds with novel structural features are incorporated into the training set for improving the quality of the models so developed.

### **3.2.2. Lead Discovery by Ligand-Based *in silico* Screening: From Dry Selection to Wet Evaluation.**

One of the main objectives of the approach developed here is the selection of *subsystems* from a large group of chemical-organic compounds. A subsystem is understood, in general, as a number of compounds formed by a significant variation in a given parent structure, which is referred to as the *lead compound* [46]. The QSAR equations found by using the **TOMOCOMD-CARDD** approach recognize some structural



patterns which are not related to the common patterns that appear in active compounds already predicted by the models. This strategy permits the identification of novel lead compounds having the desired activity, later they need to be synthesized, then tested for the pharmacological activity and they finally need to pass the toxicological, pharmacodynamical and pharmaceutical tests.

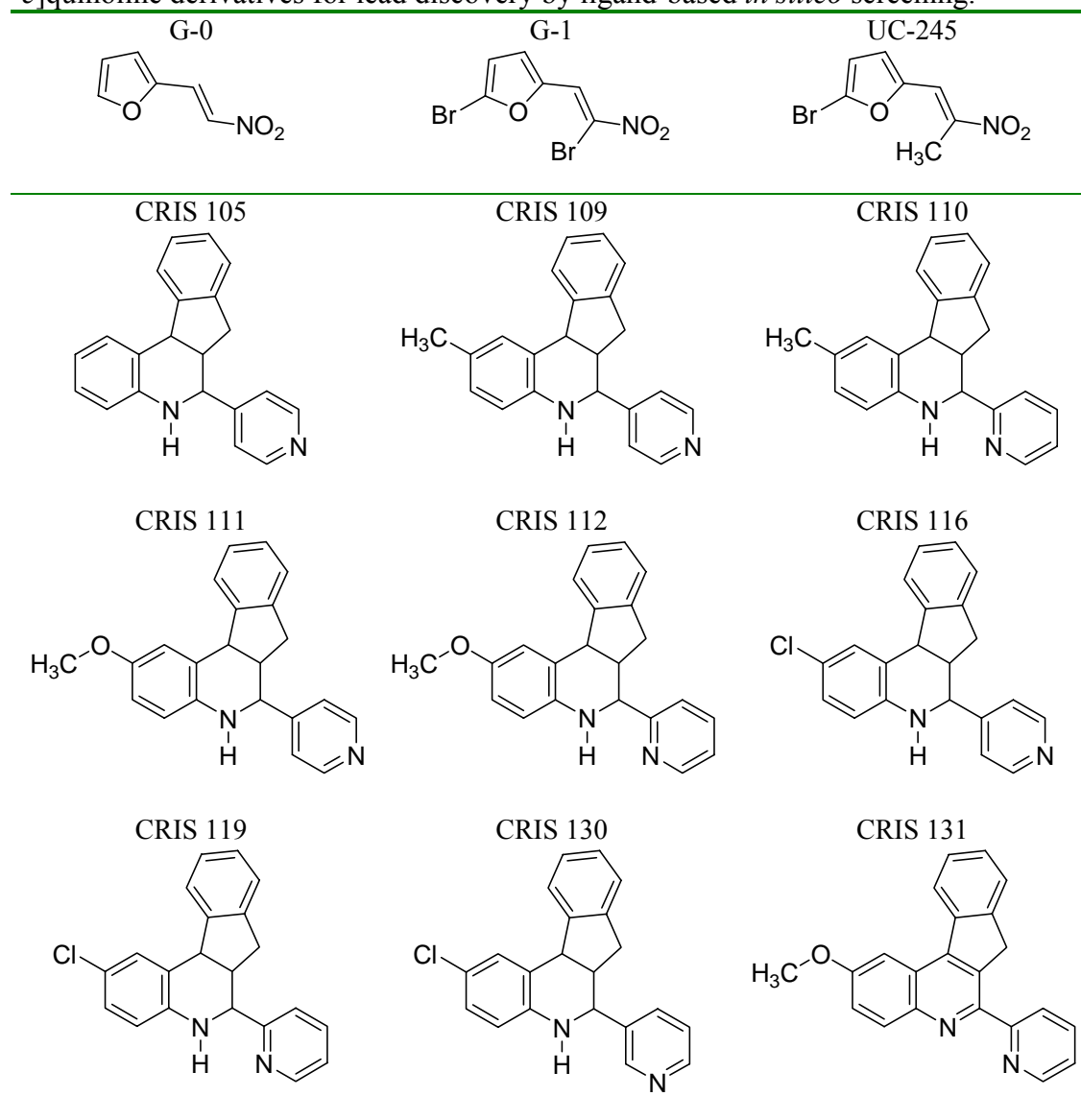
In order to test the potential of **TOMOCOMD-CARDD** method and LDA for detecting novel antiprotozoan leads, we predicted the biological activity of all the chemicals contained in our ‘in-house’ collection of nitrovinyl-furans and pyridinyl substituted 7*H*-indeno[2.1-*c*]quinoline derivatives which were provided by two of our synthesis research teams [113,115-122]. The structures of these compounds are presented in Figure 2.

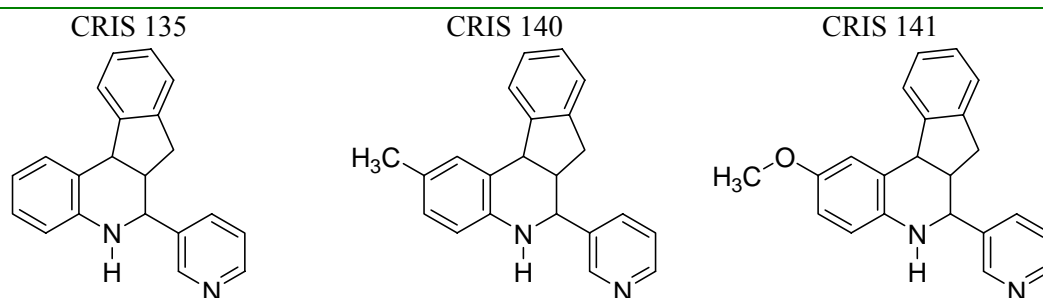
All these compounds were initially evaluated with the QSAR models **12** and **13** and then they were evaluated *in vitro*, in order to corroborate the predictions against *Tv*. The results for the classification and the  $\Delta P\%$  values of the compounds in these series are summarized in Table 8. At the same time, this table also depicts the *in vitro* antitrichomonal activity of these twenty-one compounds on *Tv*.

In general, it was observed a good coincidence between the theoretical predictions and the observed activity for both active and inactive compounds. Our trained LDA-based QSAR models (Eq. **12** and Eq.**13**) successfully classified 20 out of 21 compounds yielding (both) an accuracy of the 95.24%. In these experiments, compounds G-0, G-1, UC-245 and CRIS-148 exhibited pronounced cytotoxic activities at the concentrations of 100  $\mu\text{g/ml}$  at 24h and 48h, almost all of them showed cytotoxic activity of 100%. Compounds G-1 and UC-245 maintained a good trichomonacidal (cytotoxic) activity at

10 $\mu$ g/ml, although only G-1 maintained a high level of percentage of reduction of *T.v* at concentrations of 10mg/mL at 24h and 48h of 100% in both periods of time. On the contrary, chemicals from CRIS-105 to CRIS-153 but CRIS-148 resulted to be inactive at all assayed concentrations; coinciding with model predictions. It is remarkable that these compounds did not show toxic activity in macrophages cultivations at these concentrations (see Table 8).

**Figure 2.** Structures of nitrovinyl-furans and pyridinyl substituted 7*H*-indeno[2,1-*c*]quinoline derivatives for lead discovery by ligand-based *in silico* screening.





**Table 8.** Results of the computational evaluation using LDA-based QSAR models and percentages of cytostatic and/or cytotoxic activity [brackets] for the three concentrations assayed *in vitro* against *Tv*.

Compound*	Theoretical results				<i>in vitro</i> activity ( $\mu\text{g/ml}$ ) <sup>f</sup>						
	Class <sup>a</sup>	$\Delta\text{P}\%$ <sup>b</sup>	Class <sup>c</sup>	$\Delta\text{P}\%$ <sup>d</sup>	Class <sup>e</sup>	%CA <sub>24h</sub> [%C <sub>24h</sub> ]			%CA <sub>48h</sub> [%C <sub>48h</sub> ]		
						100	10	1	100	10	1
G-0	+	61.11	+	46.27	+	[100]	50.8	24.5	[100]	12.08	0.8
G-1	+	77.34	+	99.72	+	[100]	[100]	22.42	[100]	[100]	0
UC-245	+	62.08	+	97.80	+	[100]	33.33	0	[100]	[94.18]	6.84
CRIS 105	-	-93.16	-	-95.33	-	41.18	12.53	1.79	14.75	4.26	0.00
CRIS 109	-	-94.23	-	-94.96	-	90.28	0.00	0.00	93.77	0.00	0.00
CRIS 110	-	-93.19	-	-96.38	-	42.28	4.86	0.00	23.61	0.00	0.00
CRIS 111	-	-87.83	-	-53.01	-	90.28	13.55	0.00	84.90	0.00	0.00
CRIS 112	-	-85.74	-	-64.09	-	84.65	28.39	0.26	85.90	0.00	0.00
CRIS 116	-	-70.66	-	-77.55	-	20.23	2.64	0.00	3.82	0.00	0.00
CRIS 119	-	-66.14	-	-83.47	-	40.18	0.00	0.00	25.19	0.00	0.00
CRIS 130	-	-70.58	-	-79.34	-	74.19	40.18	18.48	95.11	5.34	1.53
CRIS 131	-	-65.10	-	-48.22	-	26.69	0.00	0.00	0.00	0.00	0.00
CRIS 135	-	-93.14	-	-95.73	-	88.86	7.92	0.00	83.59	1.53	0.00
CRIS 140	-	-94.21	-	-95.40	-	17.12	3.30	0.00	13.15	0.00	0.00
CRIS 141	-	-87.79	-	-56.27	-	68.17	0.00	0.00	86.81	0.00	0.00
CRIS 142	-	-79.28	-	-94.70	-	49.43	0.00	17.11	0.00	0.00	0.00
CRIS 143	-	-82.23	-	-94.30	-	59.09	3.42	11.43	43.27	0.00	0.00
CRIS 147	-	-82.28	-	-94.21	-	48.67	0.00	0.00	13.45	0.00	0.00
CRIS 148	-	-82.32	-	-94.40	+	[96.58]	2.66	5.70	[97.93]	0.00	0.00
CRIS 149	-	-84.88	-	-93.98	-	8.58	0.00	0.00	0.00	0.00	0.00
CRIS 153	-	-28.13	-	-74.22	-	0.00	0.00	0.00	0.00	0.00	0.00
MTZ	+	69.08	+	91.09	+	99.63	99.18	98.19	100	99.72	98.79

\*The molecular structures of the compounds represented with codes are shown in Figure 2. <sup>a,c</sup>*In silico* classification obtained from models Eq. 12 and Eq. 13 using non-stochastic and stochastic bond-type quadratic, respectively. <sup>b,d</sup>Results for the classification of compounds obtained from models Eq. 12 and Eq. 13, correspondingly:  $\Delta\text{P}\% = [\text{P}(\text{Active}) - \text{P}(\text{Inactive})] \times 100$ . <sup>e</sup>Observed (experimental activity) classification against *Tv*. <sup>f</sup>Pharmacological activity of each tested compound, which was added to the cultures at doses of 100, 10 and 1  $\mu\text{g/ml}$ : %CA<sub>#</sub> = Cytostatic activity (<sub>24 or 48 hours</sub>) and [%C<sub>#</sub>] = Cytocidal activity (<sub>24 or 48 hours</sub>). MTZ = Metronidazole (concentrations for MTZ were 2, 1 and 0.5 mg/ml, respectively).

These last results can be considered as a promising starting point for the future design and refinement of novel compounds with higher antitrichomonal activity with low toxicity. Although compounds G-1, UC-245, G-0 and CRIS-148 were active at higher doses than metronidazole, MTZ (reference drug), this result leaves a door open to a virtual variational study of the structure of these compounds in order to improve their activity. Besides, these chemicals can be taken as *hits*, which are amenable for further chemical optimization in order to derive the appropriate combination of potency, pharmacokinetic properties, toxicity etc., as well as good activity in animal models.

### 3.3. Biological *in vivo* assays of G-1.

Wistar ovariectomized rats were used in the *in vivo* experiment of G-1. The results are shown in Table 9. The 95% of the rats were infected to the beginning of the experiment, higher than previous reports [106]. The product showed 100% of effectiveness to the concentration of 0,125%. With two treatments a reduction of 50% in the infected animals was observed and with the application of the third treatment infected animals were not observed. The MTZ was also used as control in this experiment. With the application of the second treatment of MTZ infected rats were not observed. The infection controls and excipient remained infected until the end of the experiment.

**Table 9.** Results of the activity of G-1 in Wistar rats inoculated with *Tv*.

Group	Treatment	Diagnostic / % of infection						
		1	2	3	4	5	6	7
I	G-1	100	50	10	0	0	0	0
II	Migliol	80	80	80	80	80	80	80
III	Control without treatment	100	100	100	100	100	100	100
IV	Metronidazol	100	20	0	0	0	0	0
	% of total infection	95						

#### 4. CONCLUDING REMARKS

Combined features of **TOMOCOMD-CARDD** MDs and LDA techniques are able of generating polynomial QSAR models with rather statistical robustness which are potent ligand-based *virtual* and *in silico* screening tools capable of identifying drugs with a broader spectrum of antiprotozoan activity. Mainly, these *biosilico* models permit us to classify new ‘physical’ or ‘virtual’ chemicals as active or inactive ones in the chemotherapy of the trichomoniasis, and they will contribute to a more rational discovery of new lead compounds with antitrichomonal activity. In fact, this report showed that following this procedure four new chemicals with potentialities (at least in *in vitro* assays) in antitrichomonal therapeutics were found. One of them (G-1) was even a promissory drug-like compound in the treatment of this disease in more complex living systems (rats). All these compounds possess structural features not seen in known trichomonacidal and thus can serve as excellent leads for further optimisation of antitrichomonal activity. The identification of this new family, making use of the **TOMOCOMD-CARDD** approach, constitutes an example of how this rational computer-aided design method can help to reduce cost, and to increase the rate in which NCEs progress through the pipeline.

#### 5. REFERENCES AND NOTES

- [1] Heine, P., and McGregor, J. A. *Trichomonas vaginalis*: a reemerging pathogen. Clin. Obstet. Gynecol. 1993, 36, 137–144.
- [2] García, S., and Bruckner, D. A. Diagnostic Medical Parasitology. 2<sup>nd</sup> ed. Washington (D.C): American Society for Microbiology;1993:84-91.

- [3] Rein, M.F. Trichomoniasis. En: Goldsmith R, Heyneman D, editores. Parasitología y Medicina Tropical. 1<sup>ra</sup> ed. Santafé de Bogotá: El Manual Moderno; 1995.
- [4] Gram, I.T., Macaluso, M., Churchill, J., and Stalsberg, H. *Trichomonas vaginalis* (TV) and human papillomavirus (HPV) infection and the incidence of cervical intraepithelial neoplasia (CIN) grade III. *Cancer Causes and Control*. 1992, 3, 231-236.
- [5] Zhang, Z.F., Begg, C.B. Is *Trichomonas vaginalis* a cause of cervical neoplasia? Results from a combines analyses of 24 studies. *Int. J. Epidemiol*. 1994, 23, 682-690.
- [6] Viikki, M., Pukkala, E., Nieminen, P., and Hakama, M. Gynaecological infections as risk determinants of subsequent cervical neoplasia. *Acta Oncol*. 2000, 39, 71-75.
- [7] Kharsany, B.M., Hoosen, A.A., Moodley J., Bagaratee J., and Gouws, E. The association between sexually transmitted pathogens and cervical intra-epithelial neoplasia in a developing community. *Genitourin. Med*. 1993, 69, 357–360.
- [8] Cates, W., Joesoef, M.R., and Goldman, M.B. Atypical pelvic inflammatory disease: can we identify clinical predictors? *Am. J. Obstet. Gynecol*. 1993, 169, 341–346.
- [9] Grodstein, F., Goldman, M.B., and Cramer, D.W. Relation of tubal infertility to history of sexually transmitted diseases. *Am. J. Epidemiol*. 1993, 137, 577–584.
- [10] Soper, D.E., Bump, R.C., and Hurt, W.G. Bacterial vaginosis and trichomoniasis vaginitis are risk factors for cuff cellulitis after abdominal hysterectomy. *Am. J. Obstet. Gynecol*. 1990, 163, 1016–1023

- [11] Cotch, M.F. Vaginal infections and prematurity study group. Carriage of *Trichomonas vaginalis* (Tv) is associated with adverse pregnancy outcome. In: Program and abstracts of the 30th Interscience Conference on Antimicrobial Agents and Chemotherapy. American Society for Microbiology, Washington D.C., 1990, abstr. 681.
- [12] Minkoff, H., Grunebaum, A.N., Schwarz, R.H., Feldman, J., Cummings, M., Crombleholme, W., Clark, L., Pringle, G., and McCormack, W.M. Risk factors for prematurity and premature rupture of membranes: a prospective study of the vaginal flora in pregnancy. *Am. J. Obstet. Gynecol.* 1984, 150, 965–972.
- [13] Fowler, K.B., and Pass, R.F. Sexually transmitted diseases in mothers of neonates with congenital cytomegalovirus infection. *J. Infect. Dis.* 1991, 164, 259–264.
- [14] Laga, M., Manoka, A., Kivuvu, M., Malele, B., Tuliza, M., Nzila, N., Goeman, J., Behets, F., Batter, V., Alary, M., Heyward, W.L., Ryder, R.W., and Piot, P. Non-ulcerative sexually transmitted diseases as risk factors for HIV-1 transmission in women: results from a cohort study. *AIDS.* 1993, 7, 95-102.
- [15] Sorvillo, F., and Kerndt, P. *Trichomonas vaginalis* and amplification of HIV-1 transmission. *Lancet.* 1998, 351, 213-214.
- [16] Maeda, K., Osata, T., and Umezawa, H. A new antibiotic, azomycin. *J. Antibiot.* 1953, 6(Suppl. A), 182.
- [17] Cosar, C., and Julou, L. Activite' de 19 (hydroxy-2-e'thyl)-1-me'thyl-2-nitro-5 imidazole (8.823 R.P.) vis-a`-vis des infections expe'rimentales *Trichomonas vaginalis*. *Ann. Inst. Pasteur.* 1959, 96, 238–24.

- [18] 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.
- [19] Garcia-Leverde, A., and de Bonila, L. Clinical trials with metronidazole in human balantidiasis. *Am. J. Trop. Med. Hyg.* 1975, 24, 781–783.
- [20] Powell, S. J., Macleod, L., Wilmot, A. J., and Elsdon-Dew, R. Metronidazole in amoebic dysentery and amoebic liver abscess. *Lancet.* 1966, **ii**, 1329–1331.
- [21] Scheider, J. Traitment de la giardiase (lamblia) par le métronidazole. *Bull. Soc. Pathol. Exot.* 1961, 54, 84–93.
- [22] Townson, S. M., Boreham, P. F. L., Upcroft, P., Upcroft, J. A. Resistance to the nitroheterocyclic drugs. *Acta Trop.* 1994, 56.173–194.
- [23] Knight, R. The chemotherapy of amoebiasis. *J Antimicrob Chemother.* 1980, **6**, 577–593.
- [24] World Health Organization. An overview of selected curable sexually transmitted diseases, *In* Global program on AIDS. World Health Organization, Geneva, Switzerland. 1995, 2-27.
- [25] Yarlett, N., Yarlett, N.C., Lloyd, D. Ferredoxin-dependent reduction of nitroimidazole derivatives in drug-resistant and susceptible strains of *Trichomonas vaginalis*. *Biochem. Pharmacol.* 1986, 35, 1703–1708.
- [26] Tocher, J. H., and Edwards, D. I. Evidence for the direct interaction of reduced metronidazole derivatives with DNA bases. *Biochem. Pharmacol.* 48, 1089-1094.



- [27] Nielsen M H. In vitro effect of metronidazole on the ultrastructure of *Trichomonas vaginalis*. Acta Pathol. Microbiol. Scand. Sect B. 1976, 84, 93–100.
- [28] Arnold, M. Beobachtungen und Probleme bei der Behandlung der *Trichomonas vaginalis*. Ther Umsch. 1966, 23, 356–359.
- [29] Aure, J.C., and Gjonnaess H. Metronidazole treatment of trichomonal vaginitis. A comparison of cure rates in 1961 and 1967. Acta Obstet. Gynecol. Scand. 1969, 48, 440–445.
- [30] de Carneri, I. In A. Corradetti (ed.), Proceedings of the First International Congress of Parasitology, vol. 1. Pergamon Press, New York. 1966. pp. 366-367.
- [31] de Carneri, I., Baldi, G. F., Giannone, R., Passalia. S. Osservazioni preliminari su ceppi di trichomonas vaginalis naturalmente resistenti al metronidazolo. Arch. Ostet. Ginecol. 1963, 68, 422–432.
- [32] Diddle, A.W. *Trichomonas vaginalis*: resistance to metronidazole. Am. J. Obstet. Gynecol. 1967, 98, 583–585.
- [33] Giannone, T. Rilievi sull'aumento della resistenza al metronidazolo dei ceppi di *Trichomonas vaginalis* in Lombardia. Minerva Ginecol. 1972, 24, 354–355.
- [34] Kurnatowska, A. Metronidazole resistance of *Trichomonas vaginalis* Donné. Wiad Parazytol. 1969, 15, 399–401.
- [35] Robinson, S. C. Trichomonal vaginitis resistant to metronidazole. Can. Med. Assoc. J. 1962, **86**, 665.
- [36] Korner, B., and Jensen, H.K. Sensitivity of *Trichomonas vaginalis* to metronidazole, tinidazole, and nifuratel in vitro. Br J Vener Dis. 1976, 52, 404–408.

- [37] McFadzean, J.A., Pugh, L.M., Squires, S.L., Whelan, J.P. Further observations on strain sensitivity of *Trichomonas vaginalis* to metronidazole. Br. J. Vener. Dis. 1969, 45, 161–162.
- [38] Roe, F.J. Metronidazole: review of uses and toxicity. J. Antimicrob. Chemother. 1977, 3, 205–212.
- [39] Kane, P.O., McFadzean, J.A., Squires, S. Absorption and excretion of metronidazole. II. Studies on primary failures. Br J Vener Dis. 1961, 37, 276–277.
- [40] Nicol, C. S.; Evans, A. J.; McFadzean, J. A.; Squires, S. L. Lancet ii. **1966**, 441.
- [41] Meingassner, J. G., and Thurner, J. Strain of *Trichomonas vaginalis* resistant to metronidazole and other 5-nitroimidazoles. Antimicrob. Agents Chemother. 1979, 15, 254-257.
- [42] Sobel, J.D., Nyirjesy, P., and Brown, W. Tinidazole therapy for metronidazole-resistant vaginal trichomoniasis. Clin. Inf. Dis. 2001, 33, 1341-1346.
- [43] Lumsden, W.H.R., Robertson, D.H.H., Heyworth, R., and Harrison, C. Treatment failure in *Trichomonas vaginalis* vaginitis. Genitourin. Med. 1988, 64, 217.
- [44] Narcisi, E.M., Secor, W.E. In vitro effect of tinidazole and furazolidone on metronidazole-resistant *Trichomonas vaginalis*. Antimicrob. Agents Chemother. 1996, 40, 1121.
- [45] Estrada, E.; Peña, A.I.: *In silico* Studies for the Rational Discovery of Anticonvulsant Compounds. Bioorg. Med. Chem. 8, 2755, (2000).

- [46] Estrada E., Uriarte E., Montero A., Teijeira M., Santana L., De Clercq E.A.: A Novel Approach for the Virtual Screening and Rational Design of Anticancer Compounds. *J. Med. Chem.* 43, 1975, (2000).
- [47] Scott R.K.: Informatics integration: the bedrock of NCE selection. *Biosilico.* 1, 14, (2003).
- [48] Seifert M.H.J., Wolf K., Vitt D.: Virtual highthroughput *in silico* screening *Biosilico.* 1, 143, (2003).
- [49] Marrero-Ponce Y, Romero V (2002) **TOMOCOMD** software. Central University of Las Villas. **TOMOCOMD (TO**logical **MO**lecular **COM**puter **D**esign) for Windows, version 1.0 is a preliminary experimental version; in future a professional version will be obtained upon request to Y. Marrero: yovanimp@qf.uclv.edu.cu or ymarrero77@yahoo.es
- [50] 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.* 8, 687, (2003).
- [51] 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.* 44, 2010, (2004).
- [52] 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.* 12, 6351, (2004).

- [53] 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*. 9, 1100, (2004).
- [54] Marrero-Ponce Y., González-Díaz H., Romero-Zaldivar 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  $\sigma$ -Receptor Antagonist Activities. *Bioorg. Med. Chem.* 12, 5331, (2004).
- [55] 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.* 4, 512, (2003).
- [56] Marrero-Ponce Y., Cabrera M.A., Romero, V., González, D.H., Torrens F.A.: A New Topological Descriptors Based Model for Predicting Intestinal Epithelial Transport of Drugs in Caco-2 Cell Culture. *J. Pharm. Pharmaceut. Sci.* 7, 186, (2004).
- [57] 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.* 4, 124, (2005).
- [58] Marrero-Ponce Y., Castillo-Garit J.A., Olazabal E., Serrano H.S., Morales A., Castañedo N., Ibarra-Velarde F., Huesca-Guillen A., Jorge E., Sánchez A.M.,

- Torrens F., Castro E.A.: Atom, Atom-Type and Total Molecular Linear Indices as a Promising Approach for Bioorganic & Medicinal Chemistry: Theoretical and Experimental Assessment of a Novel Method for Virtual Screening and Rational Design of New Lead Anthelmintic. *Bioorg. Med. Chem.* 13, 1005, (2005).
- [59] 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.* 18, 615, (2004).
- [60] 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)* 717, 67, (2005).
- [61] Marrero-Ponce Y., Montero-Torres A., Romero-Zaldivar C., Iyarreta-Veitía I., Mayón Pérez M., García Sánchez R.: 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.* 13, 1293, (2005).
- [62] 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.* 13, 2881, (2005).

- [63] Marrero-Ponce Y., Medina-Marrero R., Martinez Y., Torrens F., Romero-Zaldivar V., Castro E.A. 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. Mod.* 2006; 12: 255–71.
- [64] Marrero-Ponce Y., Nodarse D., González-Díaz H., Ramos de Armas R., Romero-Zaldivar V., Torrens F., Castro E.: Nucleic Acid Quadratic Indices of the "Macromolecular Graph's Nucleotides Adjacency Matrix". Modeling of Footprints after the Interaction of Paromomycin with the HIV-1  $\Psi$ -RNA Packaging Region. *Int. J. Mol. Sci.* 5, 276, (2004).
- [65] Marrero-Ponce Y., Castillo-Garit J.A., Nodarse D.: Linear Indices of the "Macromolecular Graph's Nucleotides Adjacency Matrix" as a Promising Approach for Bioinformatics Studies. 1. Prediction of Paromomycin's Affinity Constant with HIV-1  $\Psi$ - RNA Packaging Region. *Bioorg. Med. Chem.* 13, 3397, (2005).
- [66] Marrero-Ponce Y., Medina R., Castro E. A., de Armas R., González H., Romero V., Torrens F.: Protein Quadratic Indices of the "Macromolecular Pseudograph's  $\alpha$ -Carbon Atom Adjacency Matrix". 1. Prediction of Arc Repressor Alanine-mutant's Stability. *Molecules.* 9, 1124, (2004).

- [67] Marrero-Ponce Y., Medina-Marrero R., Castillo-Garit J.A., Romero-Zaldivar V., Torrens F., Castro E.A.: Protein Linear Indices of the “Macromolecular Pseudograph’s  $\alpha$ -Carbon Atom Adjacency Matrix” in Bioinformatics. 1. Prediction of Protein Stability Effects of a Complete Set of Alanine Substitutions in Arc Repressor. *Bioorg. Med. Chem.* 13, 3003, (2005).
- [68] Marrero-Ponce Y., Torrens F. “Bond-Based Global and Local (Bond and Bond-Type) Quadratic Indices and Their Applications to Computer-Aided Molecular Design. 1. QSPR Studies of Octane Isomers”. *J. Comp-Aided Mol. Des.* Accepted for Publication.
- [69] Rouvray, D.H., in Balaban A.T. (Ed), In *Chemical Applications of Graph Theory*, Academic Press, London, 1976, pp 180-181.
- [70] Trinajstić, N. *Chemical Graph Theory*, CRC Press, Boca Raton, FL, 1983, 2nd ed., 1992, pp 32-33.
- [71] Estrada, E., Molina, E., Novel Local (Fragment-Based) Topological Molecular Descriptors for QSPR/QSAR and Molecular Design. *J. Mol. Graphics Mod.*, 20 (2001) 54.
- [72] Estrada, E. Edge adjacency relationships and a novel topological index related to molecular volume. *J. Chem. Inf. Comput. Sci.* 1995; 35(1): 31-3.
- [73] Estrada, E.; Ramírez, A. Edge adjacency relationships and molecular topographic descriptors. Definition and QSAR applications. *J. Chem. Inf. Comput. Sci.* 1996; 36(4): 837-43.

- [74] Estrada, E., J. Spectral moments of the edge adjacency matrix in molecular graphs, 1. Definition and applications to the prediction of physical properties of alkanes. *J. Chem. Inf. Comput. Sci.* 1996; 36(4): 844-49.
- [75] Estrada, E., Guevara, N., Gutman, I., Extension of Edge Connectivity Index. Relationships to Line Graph Indices and QSPR Application, *J. Chem. Inf. Comput. Sci.*, 38 (1998) 428.
- [76] Estrada, E., Edge-Connectivity Indices in QSPR/QSAR Studies. 2. Accounting for Long-Range Bond Contributions. *J. Chem. Inf. Comput. Sci.*, 39 (1999) 1042.
- [77] Todeschini, R., Consonni, V. (Eds.). *Handbook of Molecular Descriptors*, Wiley-VCH, Weinheim (Germany), 2000.
- [78] Edwards, C.H., Penney, D.E. *Elementary Linear Algebra*, Prentice-Hall, Englewood Cliffs, New Jersey, USA, 1988.
- [79] Estrada, E., Vilar, S., Uriarte, E., Gutierrez, Y., In Silico Studies Toward the Discovery of New Anti-HIV Nucleoside Compounds with the Use of TOPS-MODE and 2D/3D Connectivity Indices. 1. Pyrimidyl Derivatives. *J. Chem. Inf. Comput. Sci.*, 42 (2002) 1194.
- [80] Randić, M., Correlation of Enthalpy of Octanes with Orthogonal Connectivity Indices. *J. Mol. Struct. (Theochem)* **1991**, 233, 45.
- [81] Estrada, E., Peña, A., García-Domenech, R.J., Designing Sedative/Hypnotic Compounds from a Novel Substructural Graph-Theoretical Approach. *J. Comput.-Aided Mol. Des.*, 12 (1998) 583.
- [82] Potapov, V.M. *Stereochemistry*, Mir, Moscow, 1978.



- [83] Wang, R., Gao, Y., Lai, L. Calculating partition coefficient by atom-additive method. *Perspect. Drug Discov. Des.*, 19 (2000) 47.
- [84] Ertl, P., Rohde, B., Selzer, P. Fast calculation of molecular polar surface area as a sum of fragment-based contributions and its application to the prediction of drug transport properties. *J. Med. Chem.* 2000; 43: 3714-7.
- [85] Ghose, A.K., Crippen, G.M. Atomic physicochemical parameters for three-dimensional-structure-directed quantitative structure-activity relationships. 2. Modeling dispersive and hydrophobic interactions. *J. Chem. Inf. Comput. Sci.* 1987; 27: 21-35.
- [86] Millar, K.J. Additivity methods in molecular polarizability. *J. Am. Chem. Soc.* 1990; 112: 8533-42.
- [87] Gasteiger, J., Marsilli, M.A. A new model for calculating atomic charge in molecules. *Tetrahedron Lett.* 1978; 34: 3181-4.
- [88] Pauling, L. *The Nature of Chemical Bond*, Cornell University Press, Ithaca (New York), 1939, pp. 2-60.
- [89] Browder, A. *Mathematical Analysis. An Introduction*, Springer-Verlag, New York, Inc. 1996, pp 176-296.
- [90] Axler, S. *Linear Algebra Done Right*. Springer-Verlag, New York, 1996, pp 37-70.
- [91] Walker, P.D., Mezey, P.G., *Molecular Electron Density Lego Approach to Molecule Building* . *J. Am. Chem. Soc.*, 115 (1993) 12423.
- [92] Klein, D.J. Graph theoretically formulated electronic-structure theory. *Internet Electron. J. Mol. Des.* 2003; 2: 814-34.

- [93] Dmitriev I.S. *Molecules Without Chemical Bonds*, Mir publishers: Moscow, (1981).
- [94] Todeschini, R.; Gramatica, P. New 3D molecular descriptors: the WHIM theory and QSAR applications. *Perspect. Drug Dis. Des.* **1998**, 9-11, 355–380.
- [95] 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.
- [96] Kier, L. B.; Hall, L. H. *Molecular Connectivity in Structure–Activity Analysis*; Research Studies Press: Letchworth, U. K, 1986.
- [97] M. Negwer, (Ed), *Organic-Chemical Drugs and their Synonyms*; Akademie-Verlag, Berlin, 1987.
- [98] Chapman & Hall. *The Merck Index*, ver. 12:3, (1999).
- [99] van de Waterbeemd H.: In *Chemometric Methods in Molecular Design*, van Waterbeemd, H., Ed.; VCH Publishers: Weinheim, pp. 265-288, (1995).
- [100] STATISTICA Software (data analysis software system), vs 6.0; StatSoft Inc., 2001. [www.statsoft.com](http://www.statsoft.com).
- [101] Estrada E., Patlewicz G.: On the Usefulness of Graph-theoretic Descriptors in Predicting Theoretical Parameters. Phototoxicity of Polycyclic Aromatic Hydrocarbons (PAHs). *Croat. Chem. Acta.* 77, 203, (2004).
- [102] Wold S, Erikson L. In *Chemometric Methods in Molecular Design*, van Waterbeemd, H., Ed.; VCH Publishers: Weinheim, pp. 309-318, (1995).

- [103] Baldi P., Brunak S., Chauvin Y., Andersen C.A., Nielsen H. Assessing the Accuracy of Prediction Algorithms for Classification: an Overview. *Bioinformatics*. 16, 412, (2000).
- [104] 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. *Arch. Pharm. (Weinheim, Ger.)*, **2005**, 338, 1.
- [105] 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. Synthesis of 4,6-dimethyl-5-nitro(amino)-2-pyridylquinolines and their antiparasitic activities. *Arch. Pharm. (Weinheim)*. 2004; 337:127-32.
- [106] Meinsgassner JG, Georgopolus A, Patoschka M. Intravaginale infektionen der Ratte mit *Trichomonas vaginalis* und *Candida albicans*. Ein Modell zur experimentellen Chemotherapie. *Trop.Parasitol.*1975;26:395-8.
- [107] Gálvez J., García,R., Salabert M.T., Soler R.: Charge Indexes. New Topological Descriptors. *J. Chem. Inf. Comput. Sci.* 34, 520, (1994).
- [108] Johnson R.A., Wichern, D.W.: *Applied Multivariate Statistical Analysis*. Prentice-Hall: Englewood Cliffs (NJ), (1988).
- [109] Xu J., Hagler A.: *Chemoinformatics and Drug Discovery*. *Molecules*. 7, 566, (2002).
- [110] Mc Farland J.W. Gans D.J.: In *Chemometric Methods in Molecular Design*; van de Waterbeemd, H., Ed.; VCH Publishers: Weinheim, pp 295–307, 1995.

- [111] Estrada E., Uriarte E. Recent Advances on the Role of Topological Indices in Drug Discovery Research. *Curr. Med. Chem.* 8, 1573, (2001).
- [112] Alcalde, E.; Pérez, L.; Dinarés, I.; Frigola, J. *Chem. Pharm. Bull.* **1995**, 43, 493.
- [113] Ochoa, A.; Pérez, E.; Pérez, R.; Suárez, M.; Ochoa, E.; Rodríguez, H.; Gómez, A.; Muelas, S.; Nogal, R.J.J.; Martínez R.A.. Synthesis and antiprotozoan properties of new 3,5 disubstitued tetrahydro-2H-1,3,5-thiadiazine-2-thione derivatives. *Arzneim. Forsch.* 1999; 49: 764-9.
- [114] Gavini E.; Juliano C.; Mulé A.; Pirisino G.; Murineddu G.; Pinna A. Synthesis and in vitro antimicrobial properties of N-oxide derivates based on tricyclic indeno[2,1-c] pyridazine and benzo [f] cinnoline systems. *Arch. Pharm. (Weinheim)* 2000; 333: 341-6.
- [115] González, J.B; Creus, A; Marcos, R. Genetoxic evaluation of the furylethylene derivative 2-furyl-1-nitroethene in cultured human lymphocytes. *Mutation Research.*2001.497:177-84.
- [116] Castañedo, N.R; Goizueta, R.D; Gónzales, O; Peréz J.A; Gónzales, J; Silveira, E.A y col. , inventores. Procedure for the obtención of 1-(5-bromofur-2-il)-2-bromo-2-nitroeteno and it's action as micricide. European solicitud. 95. 500056. 7 / 1270. April 21. 1995a.
- [117] Castañedo, N.R; Goizueta, R.D; Gónzales, O; Peréz J.A; Gónzales, J; Silveira, E.A y col. , inventores. Procedure for the obtención of 1-(5-bromofur-2-il)-2-bromo-2-nitroeteno and it's action as micricide. US solicitud 60 / 008, 011, 1995b.Oct.27.

- [118] Castañedo, N.R; Goizueta, R.D; Gónzales, O; Pérez J.A; Gónzales, J; Silveira, E.A y col. inventores. Procedure for the obtención of 1-(5-bromofur-2-il)-2-bromo-2-nitroeteno and it's action as micricide. Certificado No – 22446. Resolución No. 2190 / 1996. ONIITEM. Oficina Nacional de Invenciones, Información Técnica y Marcas. Ciudad de La Habana. Cuba. 1996.
- [119] Castañedo, N.R; Goizueta, R.D; Gónzales, O; Pérez J.A; Gónzales, J; Silveira, E.A y col. , inventores. Procedure for the obtención of 1-(5-bromofur-2-il)-2-bromo-2-nitroeteno and it's action as micricide. Canada patent. 2, 147, 594. 1999a.
- [120] Castañedo, N.R; Goizueta, R.D; Gónzales, O; Pérez J.A; Gónzales, J; Silveira, E.A y col. inventores. Procedure for the obtención of 1-(5-bromofur-2-il)-2-bromo-2-nitroeteno and it's action as micricide. Japan patent, 2, 875, 969. 1999b
- [121] Pérez, G.M; González, J.B; Casteñedo, N.C; Creus, A and Marcos; R. In vitro genotoxicity testing of the furylethylene derivate UC-245 in human cells. *Mutagenesis*. 2004.19(1):75-80.
- [122] Kouznetsov, V.V., Ochoa, C., Puentes, Arnold R. Romero Bohórquez, Susana A. Zacchino, Maximiliano Sortino, Mahabir Gupta, Yelkaira Vázquez, Ali Bahsas and Juan Amaro-Luis, “A straightforward synthetic approach to antitumoral pyridinyl substituted 7*H*-indeno[2.1-*c*]quinoline derivatives via three-component imino Diels-Alder reaction”, *Lett. Org. Chem.*, 2006, *submitted for publication*.