ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/230019724

Discovery of Novel Trichomonacidals Using LDA-Driven QSAR Models and Bond-Based Bilinear Indices as Molecular Descriptors

ARTICLE *in* QSAR & COMBINATORIAL SCIENCE · NOVEMBER 2008

Impact Factor: 1.55 · DOI: 10.1002/qsar.200610165

CITATIONS
14

READS
32

15 AUTHORS, INCLUDING:

Oscar Miguel Rivera-Borroto
Pontifical University Catholic of Ecuador in …
22 PUBLICATIONS **75** CITATIONS
SEE PROFILE

Yovani Marrero-Ponce
Universidad Tecnológica de Bolívar (UTB), …
180 PUBLICATIONS **2,685** CITATIONS
SEE PROFILE

J.J. Nogal-Ruiz
Complutense University of Madrid
59 PUBLICATIONS **931** CITATIONS
SEE PROFILE

Francisco Torrens
University of Valencia
129 PUBLICATIONS **1,926** CITATIONS
SEE PROFILE

All in-text references underlined in blue are linked to publications on ResearchGate, letting you access and read them immediately.

Available from: Oscar Miguel Rivera-Borroto
Retrieved on: 13 January 2016

QSAR & Combinatorial Science QCS

Discovery of Novel Trichomonacidals Using LDA-Driven QSAR Models and Bond-Based Bilinear Indices as Molecular Descriptors

Oscar M. Rivera-Borroto[a,b], Yovani Marrero-Ponce[a,c,d]*, Alfredo Meneses-Marcel[a,e], José Antonio Escario[e], Alicia Gómez Barrio[e], Vicente J. Arán[f]***, Miriam A. Martins Alho[g], David Montero Pereira[e], Juan José Nogal[e], Francisco Torrens[c], Froylán Ibarra-Velarde[h], Yolanda Vera Montenegro[h], Alma Huesca-Guillén[h], Norma Rivera[h] and Christian Vogel[i]**

[a] Unit of Computer-Aided Molecular "Biosilico" Discovery and Bioinformatic Research (CAMD-BIR Unit), Faculty of Chemistry-Pharmacy, Central University of Las Villas, Santa Clara, 54830 Villa Clara, Cuba, E-mail: ymarrero77@yahoo.es, ymponce@gmail.com, yovanimp@qf.uclv.edu.cu, Phone: 53-42-281192 [or 53-42-281473] (Cuba), 963543156 (València), Fax: 53-42-281130 [or 53-42-281455] (Cuba), 963543156 (València), URL: http://www.uv.es/yoma/

[b] Center of Studies on Bioinformatics (CEI), Faculty of Mathematics, Physics and Computer Science, Central University of Las Villas, Santa Clara, 54830 Villa Clara, Cuba

[c] Institut Universitari de Ciència Molecular, Universitat de València, Edifici d'Instituts de Paterna, P. O. Box 22085, E-46071 València, Spain

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

[e] Departamento de Parasitología, Facultad de Farmacia, UCM, Pza. Ramón y Cajal s/n, 28040 Madrid, Spain

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

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

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

[i] Universität Rostock, Institut für Chemie, Abteilung für Organische Chemie, Albert-Einstein-Straße 3a, 18059 Rostock, Germany

Keywords: Bond-based bilinear indices, LDA-based QSAR model, Lead generation, TOMOCOMD-CARDD software, Trichomonacidal

Received: December 19, 2006; Accepted: December 12, 2007

DOI: 10.1002/qsar.200610165

Abstract

Few years ago, the World Health Organization estimated the number of adults with trichomoniasis at 170 million worldwide, more than the combined numbers for gonorrhea, syphilis, and chlamydia. To combat this sexually transmitted disease, Metronidazole (MTZ) has emerged, since 1959, as a powerful drug for the systematic treatment of infected patients. However, increasing resistance to MTZ, adverse effects associated to high-dose MTZ therapies and very expensive conventional technologies related to the development of new trichomonacidals necessitate novel computational methods that shorten the drug discovery pipeline. Therefore, bond-based bilinear indices, new 2-D bond-based TOMOCOMD-CARDD Molecular Descriptors (MDs), and Linear Discriminant Analysis (LDA) are combined to discover novel antitrichomonal agents. Generated models, using non-stochastic and stochastic indices, are able to classify correctly the 90.11% (93.75%) and the 87.92% (87.50%) of chemicals in the training (test) sets, respectively. In addition, they show large Matthews' correlation coefficients (C) of 0.80 (0.86) and 0.76 (0.71) for the training (test) sets, respectively. The result of predictions on the 10% *full-out* cross-validation test also evidences the quality of both models. In order to test the models' predictive power, 12 compounds, already proved against *Trichomonas vaginalis* (Tv), are screened in a simulated virtual screening experiment. As a result, they correctly classified 9 out of 12 (75.00%) and 10 out of 12 (83.33%) of the chemicals, respectively, which were the most important criteria to validate the models. Finally, in order to prove the reach of TOMOCOMD-CARDD approach and to discover new trichomonacidals, these classification functions were applied to a set of

* To whom correspondence should be addressed (contact for chem- and bioinformatics methods)
** Contact for biological assays
*** Contact for chemical methods

 WILEY InterScience DISCOVER SOMETHING GREAT © 2009 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

eight chemicals which, in turn, were synthesized and tested toward *in vitro* activity against Tv. As a result, experimental observations confirm theoretical predictions to a great extent, since it is gained a correct classification of 87.50% (7/8) of chemicals. Biological tests also show several candidates as antitrichomonals, since almost all the compounds [VAM2-(3–8)] exhibit pronounced cytocidal activities of 100% at the concentration of 100 μg/mL and at 24 h (48 h) but VAM2–2: 99.37% (100%), and it is remarkable that these compounds do not show toxic activity in macrophage assays at this concentration. The Quantitative Structure–Activity Relationship (QSAR) models presented here could significantly reduce the number of synthesized and tested compounds as well as could act as *virtual shortcuts* to new chemical entities with trichomonacidal activity.

1 Introduction

Trichomonas vaginalis (Tv) is a parasitic protozoan that is the cause of trichomoniasis, a Sexually Transmitted Disease (STD) of worldwide importance. Although Tv infection is regarded primarily as a female disease, it also occurs in men [1]. Recent estimates have suggested that Tv infections account for nearly one-third of the 15.4 million cases of STDs in the United States [2]. In 1995, the World Health Organization estimated the number of adults with trichomoniasis at 170 million worldwide, more than the combined numbers for gonorrhea, syphilis, and chlamydia [3].

Tricomonoasis encompasses symptoms such as severe inflammation [4–6]. The parasite is also known to be the main cause of vaginitis, cervicitis, and urethritis in women; it may also be responsible for prostatitis and other genitourinary syndromes in men [7, 8]. Infections with this organism have been linked to various additional pathological manifestations including cervical neoplasia [9–12], atypical pelvic inflammatory disease [13], tubal infertility [14], and so on [15–23]. Moreover, this infection can elevate the risk of acquiring human immunodeficiency virus [19, 20].

Until 1959, topical vaginal preparations available against trichomoniasis provided some symptomatic relief but were ineffective as cures [24]. In 1959, a nitroimidazole derivative of a *Streptomyces* antibiotic, azomycin, was found to be highly effective in the systemic treatment of trichomoniasis [25]. This derivative was 1-[2-hydroxyethyl]-2-methyl-5-nitroimidazole, commonly referred to as Metronidazole (MTZ).

MTZ enters the cell through diffusion [3] and is activated in the hydrogenosomes of Tv [26]. Here, the nitro group of the drug is anaerobically reduced by pyruvate-ferredoxin oxidoreductase [22, 23, 26]. This results in cytotoxic nitro radical-ion intermediates, which break the DNA strands [27, 28].

The recommended MTZ regimen results in cure rates of approximately 95% [29]. In fact, now, MTZ is the drug most widely used in the treatment of anaerobic protozoan parasitic infections caused by Tv, *Giardia duodenalis*, and *Entamoeba histolytica* [25, 30–34]. However, resistance to MTZ has been demonstrated, both in field isolates of Tv from patients refractory to treatment [35–39] and in laboratory-developed strains, obtained by exposing trichomonas to sublethal pressure of the drug either *in vitro* [40–43] or *in vivo* [44, 45].

Although literature is replete with anecdotal reports and larger patient series in which MTZ resistance seems relative and may be overcome with increasing doses of the drug [46–56], high-dose therapy with MTZ, especially when prolonged, is also associated with other important complications including pancreatitis, neutropenia, and peripheral neuropathy [57, 58].

In addition, in patients who do not respond to high-dose MTZ therapy, a variety of regimens have been evaluated for possible effectiveness, with rare or only occasional success. These include zinc sulfate, povidone-iodine douche, arsenicals, non-oxynol-9 cream, mebendazole, albendazole, furazolidone, and rifabutin [59–64]. Paromomycin was previously reported to be useful in the management of resistant trichomoniasis [65, 66]. However, local side effects were considerable and can be quite severe [65, 66].

Currently, it is clear that new trichomonacidals are needed to combat resistant Tv organisms and/or to palliate adverse effects observed in some patients toward high-dose MTZ treatments. 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 [67, 68].

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, most integration projects currently include efforts to integrate the data associated with NCE generation [69]. 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 drug chemicals) and Quantitative Structure–Activity/Property Relationship (QSAR/QSPR) studies, known as TOMOCOMD-CARDD (acronym of TOpological MOlecular COMputer Design-Computer Aided "Rational" Drug Design) [70]. This method has been developed to generate 2-D (topological), 2.5 (3-D-chiral), and 3-D (topographical and geometrical) Molecular Descriptors (MDs) based on the ap-

 © 2009 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

plication of discrete mathematics and linear algebra to chemistry. This *in silico* method has been successfully applied to the prediction of several physical, physicochemical, chemical, and biological properties of organic compounds [71 – 90].

Recently, some authors have proposed new extended local (bond and bond-type) and total (whole) MDs, based on the adjacency of edges and on quadratic, linear and bilinear maps similar to those typically defined by mathematicians in linear algebra. These researchers also proposed a new matrix representation of the molecule, in the "stochastic" adjacency of edges as well as quadratic, linear and bilinear indices derived from it. Finally, the correlation ability of the new descriptors has been tested in QSPR and QSAR studies [91 – 93].

The main objective of the present report is to use non-stochastic and stochastic bond-type bilinear indices to generate predictive Linear Discriminant Analysis (LDA)-assisted QSAR models, enabling the selection of novel drug-like compounds with antitrichomonal activity. The *in vitro* evaluation of a new series of heterocyclic compounds, with antitrichomonal activity, is also presented.

2 Materials and Methods

2.1 Theoretical Support

The basis of the extension of bilinear indices, which will be given here, is the edge-adjacency matrix considered and explicitly defined in the chemical graph-theory literature [94, 95], and rediscovered by Estrada as an important source of new MDs [96 – 101]. In this section, we will first define the nomenclature to be used in this study, then the atom-based molecular vector ($\bar{x}$) will be redefined for bond characterization, by using the same approach as previously reported [91], and finally, some new definition of bond-based non-stochastic and stochastic bilinear indices, with its peculiar mathematical properties, will be given.

2.1.1 Background in Edge-Adjacency Matrix and New Edge-Relations: Stochastic Edge-Adjacency Matrix

Let $G=(V, E)$ be a simple graph, with $V=(v_1, v_2, ..., v_n)$ and $E=(e_1, e_2, ..., e_m)$ being the vertex- and edge-sets of G, respectively. Then G represents a molecular graph having n vertices and m edges (bonds). The edge-adjacency matrix **E** of G (likewise called bond adjacency matrix, **B**) is a symmetric square matrix whose elements e_{ij} are 1 if and only if edge i is adjacent to edge j [98, 101, 102]. Two edges are adjacent if they are incident to a common vertex. This matrix corresponds to the vertex-adjacency matrix of the associated line graph. Finally, the sum of the ith row (or column) of **E** is named the edge degree of bond i, $\delta(e_i)$ [96, 99, 100, 102].

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

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

where ${}^k e_{ij}$ are the elements of the kth power of **E** and the SUM of the ith row of $\mathbf{E}^k$ is named the k-order edge degree of bond i, ${}^k\delta(e_i)$. Notice that the matrix $\mathbf{ES}^k$, defined on Eq. 1, has the property that *the sum of the elements in each row* is 1. Such an $m \times m$ matrix, with non-negative entries having this property, is called a "stochastic matrix" [103].

2.1.2 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 [74, 81, 84, 86, 104]. In a way parallel 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 numerical values, which represent a certain standard bond property (bond label). Therefore, 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 $\Re^5$, $\Re^{10}$, $\Re^{15}$, ..., $\Re^m$, respectively, where m is the dimension of the real set ($\Re^m$). This approach allows encoding organic molecules, such as 2-hydroxybut-2-enenitrile through the molecular vector $\bar{w}=[w_{\mathrm{Csp3\text{-}Csp2}}, w_{\mathrm{Csp2=Csp2}}, w_{\mathrm{Csp2\text{-}Osp3}}, w_{\mathrm{H\text{-}Osp3}}, w_{\mathrm{Csp2\text{-}Csp}}, w_{\mathrm{Csp\equiv Nsp}}]$. This vector belongs to the product space $\Re^6$.

These properties characterize each kind of bond (and bond-type) within the molecule. Diverse kinds of bond weights (w_i) can be used to encode information related to each bond in the molecule. These bond labels are chemically meaningful numbers such as standard bond distances [68, 105 – 107] and standard bond dipoles [68, 105 – 107], or even mathematical expressions involving atomic weights such as atomic log P [108], surface area contributions of polar atoms [109], atomic molar refractivities [110], atomic hybrid polarizabilities [111], Gasteiger – Marsilli atomic charges [112], atomic electronegativities in Pauling's scale [113] and so on. Here we characterized each bond with the following parameter:

$$ w = \frac{x_i}{\delta_i} + \frac{x_j}{\delta_j} \tag{2} $$

In this expression, x_i can be any standard weight of the atom i bonded to atom j. The δ_i is the vertex (atom) degree of atom i. Thus, chemical information can be codified by means of two different molecular vectors, for instance, $\bar{w}=$

 © 2009 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

$[w_1, ..., w_n]$ and $\bar{u}=[u_1, ..., u_n]$, $\bar{w} \neq \bar{u}$, which is possible if different weighting schemes are used.

In the present report, we characterized each bond with mathematical expressions (Eq. 2) involving the following parameters: atomic masses (M), van der Waals volumes (V), atomic polarizabilities (P), and atomic electronegativities (K) in Mulliken scale [114]. The values of these atomic labels are shown in Table 1. From this weighting scheme, six combinations (or twelve permutations if $\bar{w}_M$-$\bar{u}_V \neq \bar{w}_V$-$\bar{u}_M$) of molecular vectors ($\bar{w}$, $\bar{u}$;$\bar{w} \neq \bar{u}$) can be computed, namely: $\bar{w}_M$-$\bar{u}_V$, $\bar{w}_M$-$\bar{u}_P$, $\bar{w}_M$-$\bar{u}_K$, $\bar{w}_V$-$\bar{u}_P$, $\bar{w}_V$-$\bar{u}_K$, and $\bar{w}_P$-$\bar{u}_K$ [115]. Here, the symbols $\bar{w}_Y$-$\bar{u}_Z$ are used, where the subscripts Y and Z mean two atomic properties from our weighting scheme and a hyphen (−) expresses the pair. In order to illustrate the latter, an example will be described in a section below.

2.1.3 Definition of Mathematical Bilinear Forms

In mathematics, a *bilinear form* in a real vector space $\Re^m$is a mapping $b : \Re^m x\, \Re^m \rightarrow \Re$, which is linear in both arguments [116–121]. That is to say, this function satisfies the following axioms for any scalar α and any choice of vectors $\bar{v}$, $\bar{w}$, $\bar{v}_1$, $\bar{v}_2$, $\bar{w}_1$, and $\bar{w}_2$ from $\Re^m$:

i. $b(\alpha\bar{v}, \bar{w}) = b(\bar{v}, \alpha\bar{w}) = \alpha b(\bar{v}, \bar{w})$

ii. $b(\bar{v}_1 + \bar{v}_2, \bar{w}) = b(\bar{v}_1, \bar{w}) + b(\bar{v}_2, \bar{w})$

iii. $b(\bar{v}, \bar{w}_1 + \bar{w}_2) = b(\bar{v}, \bar{w}_1) + b(\bar{v}, \bar{w}_2)$

That is to say, b is bilinear if it is linear in each parameter, taken separately.

Let V be a subset of the real vector space $\Re^m/V \subset \Re^m$, and $\{\bar{e}_1, \bar{e}_2, ..., \bar{e}_m\}$ the canonical basis set of $\Re^m$. This basis set permits us to write in unambiguous form any vectors $\bar{w}$ and $\bar{u}$ of V, where: $(w^1, w^2, ..., w^m) \in V$ and $(u^1, u^2, ..., u^m) \in V$are the coordinates of the vectors $\bar{w}$ and $\bar{u}$, respectively. Namely

$$\bar{w} = \sum_{i=1}^{m} w^i e_i \quad \text{and} \quad \bar{u} = \sum_{j=1}^{m} u^j \bar{e}_j \tag{4}$$

Subsequently, applying properties i., ii., iii. (Eq. 3) we thus obtain

$$b(\bar{w}, \bar{u}) = b\left(\sum_{i=1}^{m} w^i \bar{e}_i,\ \sum_{j=1}^{m} u^j \bar{e}_j\right) = \sum_{i=1}^{m}\sum_{j=1}^{m} w^i u^j b(\bar{e}_i,\ \bar{e}_j) \tag{5}$$

If we take a_{ij} as the $m \times m$ scalars $b(\bar{e}_i,\ \bar{e}_j)$ *i.e.*

$$a_{ij} = b(\bar{e}_i,\ \bar{e}_j), \quad i = \overline{1,m} \wedge j = \overline{1,m} \tag{6}$$

Then

$$b(\bar{w}, \bar{u}) = \sum_{i,j}^{m} a_{ij} w^i u^j = \bar{w}^t \cdot \mathbf{A} \cdot \bar{u}$$

$$= \begin{bmatrix} w^1 & \dots & w^m \end{bmatrix} \cdot \begin{bmatrix} a_{11} & \dots & a_{1m} \\ \dots & \dots & \dots \\ a_{m1} & \dots & a_{mm} \end{bmatrix} \cdot \begin{bmatrix} u^1 \\ \vdots \\ u^m \end{bmatrix} \tag{7}$$

As it can be seen, the equation defined for b can be written in matrix form (see Eq. 7), where $\bar{u}$ is a column vector (an $n \times 1$ matrix) of the coordinates of $\bar{u}$ in a basis set of $\Re^m$, and $\bar{w}^t$ (a $1 \times m$ matrix) is the transpose of $\bar{w}$, where $\bar{w}$ is a column vector (an $m \times 1$ matrix) of the coordinates of $\bar{w}$ in the same basis set of $\Re^m$.

Table 1. Values of the Atomic Weights Used for Bilinear Indices Calculation.

Atomic symbol	Atomic mass	*V*d*W* volume	Sanderson 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	11.249	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

 © 2009 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

Finally, we introduce the formal definition of symmetric bilinear form. Let $\Re^m$ be a real vector space and b a bilinear function in $\Re^m$. The bilinear function b is called symmetric if $b(\bar{w}, \bar{u}) = b(\bar{u}, \bar{w}), \quad \forall \bar{w}, \bar{u} \in \Re^m$ [116–121]. Then

$$b(\bar{w}, \bar{u}) = \sum_{i,j}^{n} a_{ij} w^i u^j = \sum_{i,j}^{n} a_{ji} w^j u^i = b(\bar{u}, \bar{w}) \tag{8}$$

2.1.4 Total Non-Stochastic and Stochastic Bond-Based Bilinear Indices

If a molecule consists of m bonds (*vectors of* $\Re^m$), then the kth total bilinear indices are calculated as bilinear maps (bilinear forms) on $\Re^m$, in a canonical basis set. Specifically, the kth total non-stochastic and stochastic bond bilinear indices, $b_k(\bar{w}, \bar{u})$ and ${}^s b_k(\bar{w}, \bar{u})$, are computed from these kth non-stochastic and stochastic edge adjacency matrices, $\mathbf{E}^k$ and $\mathbf{ES}^k$, as shown in Eqs. 9 and 10, correspondingly

$$b_k(\bar{w}, \bar{u}) = \sum_{i=1}^{m} \sum_{j=1}^{m} {}^k e_{ij} w^i u^j = \bar{w}^t \cdot \mathbf{E}^k \cdot \bar{u} \tag{9}$$

$${}^s b_k(\bar{w}, \bar{u}) = \sum_{i=1}^{m} \sum_{j=1}^{m} {}^k es_{ij} w^i u^j = \bar{w}^t \cdot \mathbf{ES}^k \cdot \bar{u} \tag{10}$$

where m is the number of bonds in the molecule, and w^1, ..., w^m and u^1, ..., u^m are the coordinates of the bond-based molecular vectors $\bar{w}$and $\bar{u}$, respectively, in the canonical basis set of $\Re^m$. These coordinates will in turn coincide with the vector's components, namely, w_1, ..., w_m and u_1, ..., u_m, respectively [68, 102, 106]. Therefore, those coordinates can be considered as weights (bond labels) of the molecular graph's edge. The coefficients ${}^k e_{ij}$ and ${}^k es_{ij}$ are the elements of the kth power of the matrices $\mathbf{E}$(G) and $\mathbf{ES}$(G), correspondingly, of the molecular pseudograph. The defining equations (Eqs. 9 and 10) for $b_k(\bar{w}, \bar{u})$ and ${}^s b_k(\bar{w}, \bar{u})$, respectively, may also be written as the single matrix equation (see Eqs. 9 and 10), where $\bar{u}$ is a column vector (an $m \times 1$ matrix) of the coordinates of $\bar{u}$in the canonical basis set of $\Re^m$, and $\bar{w}^t$ is the transposed of $\bar{w}$, where $\bar{w}$ is a column vector (an $m \times 1$ matrix) of the coordinates of $\bar{w}$in the canonical basis of $\Re^m$. Here, $\mathbf{E}^k$ and $\mathbf{ES}^k$ denote the matrices of bilinear maps with regard to the natural basis set.

It should be remarked that non-stochastic and stochastic bilinear indices are symmetric and non-symmetric bilinear forms, respectively. Therefore, if in the weighting schemes, M and V are used as weights to compute theses MDs, two different sets of stochastic bilinear indices, ${}^{M-Vs} b_k^H(\bar{w}, \bar{u})$ and ${}^{V-Ms} b_k^H(\bar{w}, \bar{u})$ (because in this case $\bar{w}_M - \bar{u}_V \neq \bar{w}_V - \bar{u}_M$) can be obtained and only one group of non-stochastic bilinear indices ${}^{M-Vs} b_k^H(\bar{w}, \bar{u}) = {}^{V-Ms} b_k^H(\bar{w}, \bar{u})$ (because in this case $\bar{w}_M - \bar{u}_V = \bar{w}_V - \bar{u}_M$) can be calculated.

2.1.5 The Local Non-Stochastic and Stochastic Bond-Based Bilinear Indices

Finally, in addition to total bond-based bilinear indices, computed for the whole molecule, some local-fragmental (bond and bond-type) formalisms can be developed. These descriptors are termed as local non-stochastic and stochastic bilinear indices, $b_{kL}(\bar{w}, \bar{u})$ and ${}^s b_{kL}(\bar{w}, \bar{u})$, respectively. The definition of these descriptors is as follows:

$$b_{kL}(\bar{w}, \bar{u}) = \sum_{i=1}^{m} \sum_{j=1}^{m} {}^k e_{ijL} w^i u^j = \bar{w}^t \cdot \mathbf{E}_L^k \cdot \bar{u} \tag{11}$$

$${}^s b_{kL}(\bar{w}, \bar{u}) = \sum_{i=1}^{m} \sum_{j=1}^{m} {}^k es_{ijL} w^i u^j = \bar{w}^t \cdot \mathbf{ES}_L^k \cdot \bar{u} \tag{12}$$

where m is the number of bonds, and ${}^k e_{ijL}$ [${}^k es_{ijL}$] is the kth element of the row "i" and column "j" of the local matrix $\mathbf{E}_L^k$ [$\mathbf{ES}_L^k$]. This matrix is extracted from the $\mathbf{E}^k$ [$\mathbf{ES}^k$] matrix; it contains information referred to the edges (bonds) of the specific molecular fragments and also of the molecular environment, in k steps. The matrix $\mathbf{E}_L^k$ [$\mathbf{ES}_L^k$] with elements ${}^k e_{ijL}$ [${}^k es_{ijL}$] is defined as follows:

$$\begin{aligned} {}^k e_{ijL}[{}^k es_{ijL}] &= {}^k e_{ijL}[{}^k es_{ijL}] \text{ if both } e_i \text{ and } e_j \text{ edges (bonds) are contained within the molecular fragment} \\ &= 1/2\, {}^k e_{ijL}[{}^k es_{ijL}] \text{ if either } e_i \text{ or } e_j \text{ is contained within the molecular fragment} \\ &= 0, \text{ otherwise} \end{aligned} \tag{13}$$

It is important to highlight that the scheme above follows the spirit of the Mulliken population analysis [122]. It should be also remarked that for every partition of a molecule into Z molecular fragments there will be Z local molecular fragmental matrices. In this case, if a molecule is partitioned into Z molecular fragments, the matrices $\mathbf{E}^k$ [$\mathbf{ES}^k$] can be correspondingly partitioned into Z local matrices $\mathbf{E}_L^k$ [$\mathbf{ES}_L^k$], $L = 1, ...Z$, and the kth power of matrix $\mathbf{E}$ [$\mathbf{ES}$] is exactly the sum of the kth power of the local Z matrices. Therefore, the total (both non-stochastic and stochastic) bond-based bilinear indices are the sum of the local non-stochastic and stochastic bond-based bilinear indices, respectively, of the Z molecular fragments

$$b_k(\bar{w}, \bar{u}) = \sum_{L=1}^{Z} b_{kL}(\bar{w}, \bar{u}) \tag{14}$$

$${}^s b_k(\bar{w}, \bar{u}) = \sum_{L=1}^{Z} {}^s b_{kL}(\bar{w}, \bar{u}) \tag{15}$$

Bond and bond-type bilinear fingerprints are specific cases of local bond-based bilinear indices. The kth bond-type bilinear indices of the edge-adjacency matrix are calculated

 © 2009 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

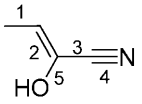
by adding the *k*th bond bilinear indices for all the bonds of the same type in the molecule. That is to say, this extension of the bond bilinear index is similar to group additive schemes, in which an index appears for each bond-type in the molecule together with its contribution based on the bond bilinear index.

In the bond-type bilinear-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 that define the bond. For all the data sets, including those with a common molecular scaffold as well as those with rather diverse structure, the *k*th fragment (bond-type) bilinear indices provide much useful information. Thus, the development of the bond-type bilinear 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 of a closely related set of structures.

It is useful to perform a calculation on a molecule to illustrate the steps in the procedure. For this, in the next section we show a representation of the computation of the non-stochastic and stochastic bilinear indices of the bond matrix (both total and local) by using a simple chemical example.

2.1.6 Sample Calculation

The bilinear indices of the bond matrix are calculated in the following way. By considering the molecule of 2-hydroxybut-2-enenitrile as a simple example, we have the following labeled molecular graph and bond-based adjacency matrices (**E** and **ES**). The second ($k=2$) and third ($k=3$) powers of these matrices as well as the bond-based molecular vectors, $\bar{w}$ and $\bar{u}$are also given



$$\mathbf{E}^0 = \mathbf{ES}^0 = \begin{bmatrix} 1 & & & & \\ & 1 & & & \\ & & 1 & & \\ & & & 1 & \\ & & & & 1 \end{bmatrix} \quad \mathbf{E}^1 = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \end{bmatrix}$$

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

$$\mathbf{ES}^1 = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0.33 & 0 & 0.33 & 0 & 0.33 \\ 0 & 0.33 & 0 & 0.33 & 0.33 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0.5 & 0.5 & 0 & 0 \end{bmatrix}$$

$$\mathbf{ES}^2 = \begin{bmatrix} 0.33 & 0 & 0.33 & 0 & 0.33 \\ 0 & 0.5 & 0.16 & 0.16 & 0.16 \\ 0.16 & 0.16 & 0.5 & 0 & 0.16 \\ 0 & 0.33 & 0 & 0.33 & 0.33 \\ 0.16 & 0.16 & 0.16 & 0.16 & 0.33 \end{bmatrix}$$

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

The molecule contains five localized bonds (corresponding to the five edges in the H-suppressed molecular graph). With these, we shall associate the five "bond orbitals" w_1, w_2, w_3, w_4, and w_5. Thus, $\bar{w}=[w_1, w_2, w_3, w_4, w_5]=[w_{(C-C)}, w_{(C=C)}, w_{(C-C)}, w_{(C\equiv N)}, w_{(C-O)}]$ and each "bond orbital" can be computed by Eq. 2 by using, for instance, the atomic electronegativity in Pauling scale (x) [113] as atomic weight (atom label)

$$w_1 = \frac{x_C}{1} + \frac{x_C}{3} = \frac{2.55}{1} + \frac{2.55}{3} = 3.4$$

$$w_2 = \frac{x_C}{3} + \frac{x_C}{4} = \frac{2.55}{3} + \frac{2.55}{4} = 1.4875$$

$$w_3 = \frac{x_C}{4} + \frac{x_C}{4} = \frac{2.55}{4} + \frac{2.55}{4} = 1.275$$

$$w_4 = \frac{x_C}{4} + \frac{x_N}{3} = \frac{2.55}{4} + \frac{3.04}{3} = 1.650833$$

$$w_5 = \frac{x_C}{4} + \frac{x_O}{1} = \frac{2.55}{4} + \frac{3.44}{1} = 4.0775$$

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

Besides, other vector, $\bar{u}$, can be calculated in the same way as $\bar{w}$, but using other property, for example, the atomic mass [114] as atomic weight (atom-label)

$$u_1 = \frac{y_C}{1} + \frac{y_C}{3} = \frac{12.01}{1} + \frac{12.01}{3} = 16.013333$$

$$u_2 = \frac{y_C}{3} + \frac{y_C}{4} = \frac{12.01}{3} + \frac{12.01}{4} = 7.005833$$

$$u_3 = \frac{y_C}{4} + \frac{y_C}{4} = \frac{12.01}{4} + \frac{12.01}{4} = 6.0050$$

$$u_4 = \frac{y_C}{4} + \frac{y_N}{3} = \frac{12.01}{4} + \frac{14.01}{3} = 7.6725$$

$$u_5 = \frac{y_C}{4} + \frac{y_O}{1} = \frac{12.01}{4} + \frac{16.00}{1} = 19.0025$$

and therefore, $\bar{u}=[16.013333, 7.005833, 6.0050, 7.6725, 19.0025]$.

Each non-stochastic or stochastic total bilinear index will have the form

 © 2009 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

$$b_k(\bar{w},\ \bar{u}) = +{}^k e_{11} w^1 u^1 + {}^k e_{12} w^1 u^2 + {}^k e_{13} w^1 u^3 + {}^k e_{14} w^1 u^4 + {}^k e_{15} w^1 u^5 + {}^k e_{21} w^2 u^1 + {}^k e_{22} w^2 u^2 + {}^k e_{23} w^2 u^3 + {}^k e_{24} w^2 u^4 + {}^k e_{25} w^2 u^5 + {}^k e_{31} w^3 u^1 + {}^k e_{32} w^3 u^2 + {}^k e_{33} w^3 u^3 + {}^k e_{34} w^3 u^4 + {}^k e_{35} w^3 u^5 + {}^k e_{41} w^4 u^1 + {}^k e_{42} w^4 u^2 + {}^k e_{43} w^4 u^3 + {}^k e_{44} w^4 u^4 + {}^k e_{45} w^4 u^5 + {}^k e_{51} w^5 u^1 + {}^k e_{52} w^5 u^2 + {}^k e_{53} w^5 u^3 + {}^k e_{54} w^5 u^4 + {}^k e_{55} w^5 u^5 = \sum_{i=1}^{5} {}^k e_{ii} w^i u^i + 2 \sum_{(i,j=1,\, i\neq j)}^{5} {}^k e_{ij} w^i u^j$$

$${}^s b_k(\bar{w},\ \bar{u}) = +{}^k es_{11} w^1 u^1 + {}^k es_{12} w^1 u^2 + {}^k es_{13} w^1 u^3 + {}^k es_{14} w^1 u^4 + {}^k es_{15} w^1 u^5 + {}^k es_{21} w^2 u^1 + {}^k es_{22} w^2 u^2 + {}^k es_{23} w^2 u^3 + {}^k es_{24} w^2 u^4 + {}^k es_{25} w^2 u^5 + {}^k es_{31} w^3 u^1 + {}^k es_{32} w^3 u^2 + {}^k es_{33} w^3 u^3 + {}^k es_{34} w^3 u^4 + {}^k es_{35} w^3 u^5 + {}^k es_{41} w^4 u^1 + {}^k es_{42} w^4 u^2 + {}^k es_{43} w^4 u^3 + {}^k es_{44} w^4 u^4 + {}^k es_{45} w^4 u^5 + {}^k es_{51} w^5 u^1 + {}^k es_{52} w^5 u^2 + {}^k es_{53} w^5 u^3 + {}^k es_{54} w^5 u^4 + {}^k es_{55} w_5 u_5 = \sum_{i=1}^{5} {}^k e_{ii} w^i u^i + \sum_{(i,j=1)}^{5} {}^k e_{ij} w^i u^j$$

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

In this way, $\mathbf{E}^k$ and $\mathbf{ES}^k$ can be seen as graph–theoretical electronic-structure models [123]. Actually, quantum chemistry starts from the fact that 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 on 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 ($k = 0, 1, 2, 3, ..., n$) through the chemical-bonding network. Therefore, the $\mathbf{E}^1$ and $\mathbf{ES}^1$ matrices consider the valence-bond electrons in one step and their power ($k = 0, 1, 2, 3, ..., n$) can be considered as an interacting – electron chemical – network model in k steps. This model can be seen as an intermediate between the quantitative quantum-mechanical Schrödinger equation and classical chemical bonding ideas [123].

On the other hand, the kth ($k = \overline{0,3}$) non-stochastic total bilinear indices can be expressed as the sum of the local (bond) bilinear indices for this molecule as follows:

$$\begin{aligned} b_0(\bar{w},\ \bar{u}) &= \sum_{L=1}^{5} b_{0L}(\bar{w},\ \bar{u}) = b_{01}(\bar{w},\ \bar{u}) + b_{02}(\bar{w},\ \bar{u}) + b_{03} \times (\bar{w},\ \bar{u}) + b_{04}(\bar{w},\ \bar{u}) + b_{05}(\bar{w},\ \bar{u}) \\ &= 54.44533 + 10.42118 + 7.656375 + 12.66602 + 77.48269 \\ &= 162.6716 \end{aligned}$$

$$\begin{aligned} b_1(\bar{w},\ \bar{u}) &= \sum_{L=1}^{5} b_{1L}(\bar{w},\ \bar{u}) = b_{11}(\bar{w},\ \bar{u}) + b_{12}(\bar{w},\ \bar{u}) + b_{13} \times (\bar{w},\ \bar{u}) + b_{14}(\bar{w},\ \bar{u}) + b_{15}(\bar{w},\ \bar{u}) \\ &= 23.81983 + 61.16852 + 43.13707 + 9.847846 + 52.77304 \\ &= 190.7463 \end{aligned}$$

$$\begin{aligned} b_2(\bar{w},\ \bar{u}) &= \sum_{L=1}^{5} b_{2L}(\bar{w},\ \bar{u}) = b_{21}(\bar{w},\ \bar{u}) + b_{22}(\bar{w},\ \bar{u}) + b_{23} \times (\bar{w},\ \bar{u}) + b_{24}(\bar{w},\ \bar{u}) + b_{25}(\bar{w},\ \bar{u}) \\ &= 139.8138 + 80.10137 + 76.67535 + 55.48246 + 304.0172 \\ &= 656.0901 \end{aligned}$$

$$\begin{aligned} b_3(\bar{w},\ \bar{u}) &= \sum_{L=1}^{5} b_{3L}(\bar{w},\ \bar{u}) = b_{31}(\bar{w},\ \bar{u}) + b_{32}(\bar{w},\ \bar{u}) + b_{33} \times (\bar{w},\ \bar{u}) + b_{34}(\bar{w},\ \bar{u}) + b_{35}(\bar{w},\ \bar{u}) \\ &= 183.0889 + 262.1182 + 207.3626 + 98.6209 + 462.33630 \\ &= 1213.527 \end{aligned}$$

The terms in the summations for calculating the total bilinear indices are the so-called local (bond) bilinear indices, which, in turn, have been written in the consecutive order of the bond labels in the graph. For instance, the non-stochastic bond bilinear indices of orders 0, 1, 2, and 3 for the bond labeled as $\underline{1}$ are 54.44533, 23.81983, 139.8138, and 183.0889, respectively.

The kth total stochastic bilinear indices values are also the sum of the kth local (bond) stochastic bilinear indices values for all bonds in the molecule

$$\begin{aligned} {}^s b_0(\bar{w},\ \bar{u}) &= \sum_{L=1}^{5} {}^s b_{0L}(\bar{w},\ \bar{u}) = {}^s b_{01}(\bar{w},\ \bar{u}) + {}^s b_{02}(\bar{w},\ \bar{u}) + {}^s b_{03} \times (\bar{w},\ \bar{u}) + {}^s b_{04}(\bar{w},\ \bar{u}) + {}^s b_{05}(\bar{w},\ \bar{u}) \\ &= 54.44533 + 10.42118 + 7.656375 + 12.66602 + 77.48269 \\ &= 162.6716 \end{aligned}$$

$$\begin{aligned} {}^s b_1(\bar{w},\ \bar{u}) &= \sum_{L=1}^{5} {}^s b_{1L}(\bar{w},\ \bar{u}) = {}^s b_{11}(\bar{w},\ \bar{u}) + {}^s b_{12}(\bar{w},\ \bar{u}) + {}^s b_{13} \times (\bar{w},\ \bar{u}) + {}^s b_{14}(\bar{w},\ \bar{u}) + {}^s b_{15}(\bar{w},\ \bar{u}) \\ &= 15.87989 + 30.70998 + 19.72389 + 6.587033 + 22.01199 \\ &= 94.91277 \end{aligned}$$

$$\begin{aligned} {}^s b_2(\bar{w},\ \bar{u}) &= \sum_{L=1}^{5} {}^s b_{2L}(\bar{w},\ \bar{u}) = {}^s b_{21}(\bar{w},\ \bar{u}) + {}^s b_{22}(\bar{w},\ \bar{u}) + {}^s b_{23} \times (\bar{w},\ \bar{u}) + {}^s b_{24}(\bar{w},\ \bar{u}) + {}^s b_{25}(\bar{w},\ \bar{u}) \\ &= 39.46198 + 14.31402 + 14.48064 + 14.93603 + 58.66773 \\ &= 141.8604 \end{aligned}$$

 © 2009 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

$$
\begin{aligned}
{}^{s}b_3(\bar{w},\ \bar{u}) = \sum_{L=1}^{5} {}^{s}b_{3L}(\bar{w},\ \bar{u}) = {}^{s}b_{31}(\bar{w},\ \bar{u}) + {}^{s}b_{32}(\bar{w},\ \bar{u}) + {}^{s}b_{33} \\
\times (\bar{w},\ \bar{u}) + {}^{s}b_{34}(\bar{w},\ \bar{u}) + {}^{s}b_{35}(\bar{w},\ \bar{u}) \\
= 23.20039 + 22.578 + 17.14819 \\
+ 13.09528 + 40.77731 \\
= 116.7992
\end{aligned}
$$

2.2 Computational Strategies

2.2.1 TOMOCOMD-CARDD Approach

TOMOCOMD is an interactive program for molecular design and bioinformatic research [70]. It consists of four subprograms; each one of them allows drawing the structures (drawing mode) and calculating molecular 2-D/3-D (calculation mode) descriptors. The modules are named Computed-Aided "Rational" Drug Design (CARDD), Computed-Aided Modeling in Protein Science (CAMPS), Computed-Aided Nucleic Acid Research (CANAR), and Computed-Aided Bio-Polymers Docking (CABPD). In the present report, salient features are outlined concerning with only one of these subprograms, CARDD, and with the calculation of non-stochastic and stochastic 2-D bond-based bilinear indices.

2.2.1.1 Work Methodology

The main steps for the application of the present method in QSAR/QSPR and drug design can be briefly summarized in the following algorithm: (i) draw the molecular pseudographs for each molecule in the data set, using the software drawing mode. This procedure is performed by a selection of the active atomic symbols belonging to the different groups in the periodic table of the elements. (ii) Use appropriated atomic properties in order to weight and differentiate the molecular bonds. In this study, the used properties are those previously proposed for the calculation of the DRAGON descriptors [113, 125, 126], *i.e.*, atomic mass (M), atomic polarizability (P), atomic Sanderson electronegativity (K), van der Waals atomic volume (V), and the atomic electronegativity in Pauling scale (G) [114]. The values of these atomic labels are shown in Table 1. In order to calculate the required weights, it has used the mathematical expression given by Eq. 2, which involves atomic weights. (iii) Compute the total and local (bond and bond-type) non-stochastic and stochastic bilinear indices. It can be carried out in the software calculation mode, where one can select the atomic properties and the descriptor family 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 linear indices or other family of MDs implemented in this program. (iv) 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 bilinear indices having, for instance, the following appearance, $A = a_0 b_0(\bar{w},\ \bar{u}) + a_1 b_1(\bar{w},\ \bar{u}) + a_2 b_2(\bar{w},\ \bar{u}) + ... + a_k b_k(\bar{w},\ \bar{u}) + c$, where A is the measured activity, $b_k(\bar{w})$ are the kth total bond-based bilinear indices, and the a_k and c are the coefficients obtained by the linear regression analysis. (v) 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. (vi) Apply the obtained LDA-based QSAR models as cheminformatic tools for identifying and/or discovering novel drugs through the ligand-based *virtual* screening procedure.

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

1) kth $\left(k = \overline{0, 15}\right)$ total non-stochastic bond-based bilinear indices, not considering and considering H-atoms in the molecular graph (G) [${}^{Y-Z}b_k(\bar{w},\ \bar{u})$ and ${}^{Y-Z}b_k^{H}(\bar{w},\ \bar{u})$], respectively.
2) kth $\left(k = \overline{0, 15}\right)$ total stochastic bond-based bilinear indices, not considering and considering H-atoms in the molecular graph (G) [${}^{Y-Zs}b_k(\bar{w},\ \bar{u})$ and ${}^{Y-Zs}b_k^{H}(\bar{w},\ \bar{u})$], respectively.
3) kth $\left(k = \overline{0, 15}\right)$ bond-type (group = heteroatoms: S, N, O) non-stochastic linear indices, not considering and considering H-atoms in the molecular graph (G) [${}^{Y-Z}b_{kLE}(\bar{w},\ \bar{u})$ and ${}^{Y-Z}b_{kLE}^{H}(\bar{w},\ \bar{u})$], respectively. These local descriptors are putative molecular charge, dipole moment, and H-bonding acceptors character.
4) kth $\left(k = \overline{0, 15}\right)$ bond-type (group = heteroatoms: S, N, O) stochastic bilinear indices not considering and considering H-atoms in the molecular graph (G) [${}^{Y\text{-}Zs}b_{kLE}$ $(\bar{w},\ \bar{u})$, and ${}^{Y-Zs}b_{kLE}^{H}(\bar{w},\bar{u})$], respectively. These local descriptors are putative molecular charge, dipole moment, and H-bonding acceptors character.

2.3 Data Management and Statistical Processing

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 as the inactive compounds set, through a random selection, also guaranteeing a great structural variability. All these chemicals were taken from the Negwer, Handbook [127] and Merck, Index

 © 2009 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

[128], where their names, synonyms, and structural formulas can be found. From these 123 chemicals 91 were chosen at random to form the training set, 40 of them being active and 51 inactive ones. The remaining subseries, consisting of 10 trichomonacidals and 22 non-trichomonacidals, were prepared as test sets for the external validation of the models. The latter 32 chemicals were not used in the development of the classification models.

The discriminant functions were obtained by using LDA [129], as implemented in the STATISTICA program [130]. 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 maximal parsimony (Occam's razor) was taken into account as the strategy for model selection. In its original form, Occam's razor states that "*Numquam ponenda est pluritas sin necesitate*," which can be translated as "Entities should not be multiplied beyond necessity" [131]. 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, 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 [131]. Therefore, we select the model with the highest statistical signification, but having as few parameters (a_k) as possible.

The quality of the models were determined by examining Wilks' λ parameter (U-statistic), square Mahalanobis' distance (D^2), Fisher's ratio (F), and the corresponding p-level [$p(F)$] as well as the percentage of good classification in the training and test sets [129]. Models with a ratio between the number of cases and variables in the equation lower than 5 were rejected.

The Wilks' λ, 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 to differentiate between the two respective groups.

The classification of cases was carried out by means of the posterior classification probabilities. Using the Mahalanobis distances to do the classification, we can now derive the probabilities. The probability that a case belongs to a particular group is basically proportional to the Mahalanobis distance from that group centroid (it is not exactly proportional because we assume a multivariate normal distribution around each centroid). Because the location of each case is computed from the prior knowledge of the values for that case on the variables in the model, these probabilities are called posterior probabilities. In summary, the posterior probability is the probability, based on our knowledge of the values of other variables, with which the respective case belongs to a particular group [130].

By using the models, any compound can be classified as either active if $\Delta P\% > 0$, being $\Delta P\% = [P(\text{active}) - P(\text{inactive})] \times 100$, or inactive, otherwise. P(active) and P(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 [132]. In addition, a Leave-Many-Out (LMO) 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 ten times on ten 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 power of the obtained models [132].

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 (predicting) sets permitted the assessment of the model [133].

2.4 Microscopic and Culture Techniques

The biological activity was tested on *Tv* JH31A #4 Ref. No. 30326 (ATCC, MD, USA), in modified Diamond medium, supplemented with equine serum and grown at 37° (5% CO_2). The compounds were added to the cultures at several concentrations (100, 10, and 1 μg/mL), after 6 h of the seeding (0 h). Viable protozoa were assessed at 24 and 48 h after incubation at 37°, by using the Neubauer chamber. MTZ (Sigma – Aldrich, SA, Spain) was used as a reference drug at concentrations of 2, 1, 0.5 μg/mL. The cytocidal and cytostatic activities were determined, by the calculation of percentages of cytocidal (%C) and cytostatic activities (%CA) in relation to controls, as previously reported [134, 135].

3 Results and Discussion

3.1 Generation and Statistical Analysis of the LDA-Driven QSAR Models

In spite of the extensive number of statistical methods to obtain classification functions, we select LDA given the simplicity of the method [129]. LDA in drug design has been extensively reported by different authors [68, 73 – 75, 79 – 87, 136, 137]. Therefore, LDA was also the technique used in the generation of discriminant functions in the current work. By means of the LDA technique implemented on the STATISTICA software [130], the following linear models were obtained, in which the total as well as local non-stochastic and stochastic bond-based bilinear indices were used as independent variables

 © 2009 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

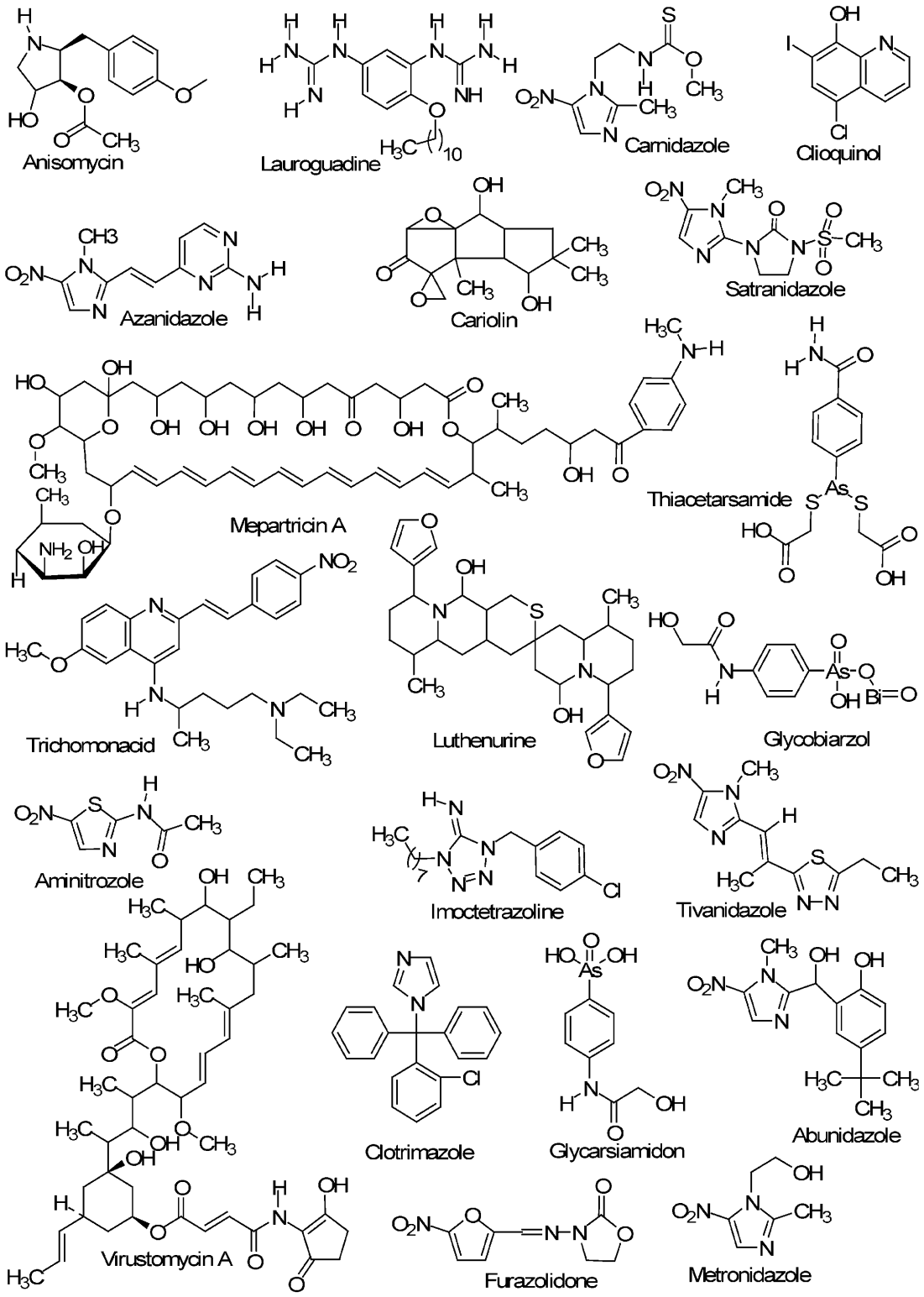


Figure 1. Random sample of families of trichomonacidal agents studied here.

 © 2009 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

$$Clasif = -3.32 - 0.02^{MP}b_1^H(\bar{w},\ \bar{u}) + 0.06^{ME}b_0^H(\bar{w},\ \bar{u}) + 0.12^{ME}b_{1L}^H(\bar{w}_{hal},\ \bar{u}_{hal}) - 0.12^{ME}b_{1L}(\bar{w}_{hal},\ \bar{u}_{hal}) - 1.40 \times 10^{-3MV}b_0(\bar{w},\ \bar{u}) - 7.59 \times 10^{-4MV}b_{0L}^H(\bar{w}_E,\ \bar{u}_E)$$
$$N = 91\ \ \lambda = 0.46\ \ D^2 = 4.51\ \ F(6,\ 84) = 15.90\ \ p < 0.0001 \quad (16)$$

$$Clasif = -5.27 + 0.23^{MP}b_{2L}^H(\bar{w}_E,\ \bar{u}_E) - 0.59^{MP}b_{3L}^H(\bar{w}_E,\ \bar{u}_E) + 0.13^{ME}b_{1L}^H((\bar{w}_E,\ \bar{u}_E) - 0.21^{VP}b_{1L}^H(\bar{w}_{hal}, \bar{u}_{hal}) + 0.43^{VP}b_{4L}^H(\bar{w}_E,\ \bar{u}_E) - 0.02^{VP}b_{3L}(\bar{w}_E,\ \bar{u}_E)$$
$$N = 91\ \ \lambda = 0.46\ \ D^2 = 4.51\ \ F(6,\ 84) = 15.90\ \ p < 0.0001 \quad (17)$$

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

As a result, model 16 was able to classify the 90.00% (36/40) of active and the 90.20% (46/51) of the inactive chemicals in the training set for a globally good classification of 90.11% (82/91). In addition, in the same learning series, model 17 correctly classified the 87.50% (35/40) of the active and the 88.24% (45/51) of the inactive compounds by yielding an accuracy of 87.92% (80/91). Similarly, Eqs. 16 and 17 showed a 93.75% (30/32) and 87.50% (28/32), respectively, of global predictability in the prediction series. Since 85.00% is considered as an acceptable threshold limit for this kind of analysis, the former results validate, to some extent, the models for their use in the ligand-based *virtual* screening procedure [138].

In Tables 2 and 3, divided into active or inactive substances, respectively, the names of all compounds in the training and test sets, together with their posterior probabilities calculated from the Mahalanobis distance by using both equations, are given.

Later, a more serious analysis was carried out by calculating most commonly used parameters in medical statistics (accuracy, sensitivity, specificity, and false positive rate) and the Matthews correlation coefficient, C. Table 4 lists these parameters for both models [133, 139]. While the sensitivity is the probability of correctly predicting a positive case, the specificity is the probability that a positive prediction be correct. The Matthews' C quantifies the strength of the linear relationship between the MDs and the classifications, and it may often provide a much more balanced evaluation of the prediction than, for instance, the percentages [133, 139]. As a result, the obtained QSAR models, Eqs. 16 and 17, showed high C of 0.80 (0.86) and 0.76 (0.71) in training (test) sets, corresponding-

Table 2. Names and classification of active compounds into training and test sets, according to both TOMOCOMD-CARDD models developed in this work.

Name	$\Delta P\%$ [a]	$\Delta P\%$ [b]	Name	$\Delta P\%$ [a]	$\Delta P\%$ [b]
Active training set					
Anisomycin	26.59	**−53.48**	Abunidazole	53.14	91.70
Virustomycin A	79.95	85.07	Imoctetrazoline	5.95	95.76
Azanidazole	86.29	98.61	Forminitrazole	87.40	98.28
Carnidazole	91.60	98.03	Chlomizol	99.14	98.68
Propenidazole	95.55	97.07	Acinitrazole	79.51	91.81
Lauroguadine	**−89.10**	**−59.23**	Moxnidazole	99.53	99.94
Mepartricin A	99.88	82.34	Isometronidazole	73.75	73.22
Metronidazole	73.75	70.96	Mertronidazole phosphate	93.26	87.87
Nifuratel	99.08	99.11	Benzoylmetronidazole	96.19	97.64
Nifuroxime	96.75	88.05	Bamnidazole	94.52	96.20
Nimorazole	92.36	69.70	Glycarsiamidon	10.63	−49.70
Secnidazole	63.12	57.21	Fexinidazole	88.90	79.94
Cariolin	**−84.67**	**−74.51**	Piperanitrozole	78.40	92.84
2-Amino-5-nitrothiazole	58.06	87.06	Gynotabs	85.30	91.91
Glycobiarzol	78.61	88.86	Pirinidazole	82.21	89.66
Clioquinol	98.19	99.77	Metronidazole hydrogen succinate	98.36	98.00
Diiodohydroxy Quinoline	98.97	−7.55	Tolamizol	92.73	98.75
Ornidazol	93.39	97.04	Thiacetarsamide	61.21	40.58
Trichomonacid	76.92	69.38	Tivanidazole	61.30	91.23
Lutenurine	**−82.33**	**−18.98**	Policresulen	**−3.11**	2.87
Active test set					
Acertarsone	**−5.17**	**−51.04**	Pentamycin	34.82	**−97.40**
Furazolidone	99.32	99.79	Azomycin	76.38	84.49
Mepartricin B	99.89	81.27	Ternidazole	70.49	60.77
Aminitrozole	79.51	91.81	Misonidazole	93.39	40.50
Clotrimazol	57.41	79.31	Satranidazole	93.53	99.73

[a] Antitrichomonal activity predicted by Eq. 16: $\Delta P\% = [P(\text{active}) - P(\text{inactive})] \times 100$.
[b] Antitrichomonal activity predicted by Eq. 17: $\Delta P\% = [P(\text{active}) - P(\text{inactive})] \times 100$.

 © 2009 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

Table 3. Names and classification of inactive compounds into training and test sets, according to both TOMOCOMD-CARDD models developed in this work.

Name	$\Delta P\%$ [a]	$\Delta P\%$ [b]	Name	$\Delta P\%$ [a]	$\Delta P\%$ [b]
Inactive training set					
Amantadine	−98.65	−99.24	Non-aferone	−96.29	−97.89
Thiacetazone	−44.75	−77.21	Rolipram	−34.28	−88.14
Cloral betaine	−97.98	−57.59	*N*-hydroxymethyl-*N*-methylurea	−81.24	−95.81
Carbavin	−76.93	−87.94	4 Chlorobenzoic acid	−72.40	**20.30**
Norantoin	−54.69	−45.89	Acetanilide	−80.34	−97.77
Orotonsan Fe	**16.86**	**59.14**	Guanazole	−89.20	−98.14
Picosulfate	**94.95**	**98.22**	Tetramin	−88.62	−99.02
Naftazone	−27.25	−81.34	Mecysteine	−80.69	−78.02
Besunide	−40.57	−43.37	Cirazoline	−70.42	−95.18
Acetazolamide	−48.56	−63.82	Methocarbamol	**71.79**	−69.69
Propamine soviet	−98.21	−98.66	Lysergide	−93.54	−87.22
RMI 11894	−98.43	−98.70	Dopamine	−84.53	−95.62
Ag 307	−89.68	−95.44	Bufeniode	−96.62	−81.71
Barbismethylii iodide	−97.55	−99.65	Celiprolol	−76.08	−89.83
Pancuronium bromide	−99.03	−91.10	Erysimin	−8.80	−45.76
Vinyl ether	−65.45	−98.54	Peruvoside	**29.47**	−31.38
Basedol	−76.80	−17.97	Amitraz	−93.60	−88.41
Carbimazole	−30.82	**79.35**	Proclonol	−94.12	**75.78**
Didym levulinate	−68.02	−97.84	Asame	−2.16	−81.28
Perchloroethane	−93.48	−96.62	KC-8973	−90.51	−97.10
Pyrantel tartrate	−89.36	−98.33	Ethydine	−14.30	−47.57
Fentanyl	−78.10	−95.12	Magnesii metioglicas	−49.18	−97.39
Petidine	−79.63	−96.06	Alibendol	−14.35	−87.68
Tenalidine tartrate	−98.56	−99.78	Diponium Bromide	−98.01	−97.89
Bamipine	−91.02	−98.23	Streptomycin	**97.81**	**84.39**
Colestipol	−95.93	−99.71			
Inactive test set					
Citenazone	−66.17	−46.26	Metriponate	−95.86	−71.35
Methenamine	−83.90	−91.56	Ciclopramine	−94.07	−93.12
Penthrichloral	−85.49	−90.23	Litracen	−97.74	−99.54
Calcium sodium ferriclate	−100	−100	Trimethylsulfonium hydroxyde	−98.96	−100
Ferroceron	−99.89	−99.88	Norgamem	−73.04	−97.19
Emodin	−19.31	−84.72	Emylcamate	−91.46	−92.70
Butanolum	−94.79	−99.47	Acetylcholine	−87.52	−99.87
Spironolactone	−95.73	−96.12	Carazolol	−71.90	−93.16
Bromcholine	−99.38	−99.16	Cefazolin	**69.40**	**99.47**
Imekhin	−99.71	−99.16	Penicillin I	−82.42	**19.66**
Diphenadione	−17.79	−96.11	Aziromycin	−76.72	−91.26

[a] Antitrichomonal activity predicted by Eq. 16 : $\Delta P\% = [P(\text{active}) - P(\text{inactive})] \times 100$.
[b] Antitrichomonal activity predicted by Eq. 17: $\Delta P\% = [P(\text{active}) - P(\text{inactive})] \times 100$.

ly, which proves the existence of a strong linear relationship, since a value of +1 implies a total linear agreement between the variables under consideration.

3.1.1 Cross-Validation Methods as the Key for QSAR Model Internal Validation

Validation is a crucial aspect of any QSAR modeling. Therefore, internal validation methods (*e.g.* cross-validation) are considered, by many authors, as an indicator or even as the ultimate proof of the stability and high-predictive power of a QSAR model. However, Golbraikh and Tropsha demonstrated that high values of leave-one-out square correlation coefficient q^2 appear to be a necessary, but not a sufficient condition for the model to have a high predictive power [140]. A more exhaustive cross-validation method can be used, in which a fraction of the data (10–20%) is left out and predicted from a model based on the remaining data. This process (LMO) is repeated until each observation has been left out at least once [140, 141].

Following these statements, a leave-ten-*fold* full-out (LMO) cross-validation procedure was carried out here. For each group of observations (10% of the whole data set, nine compounds) left out, a model was developed from the remaining 90% of the data (81 compounds). This procedure was repeated ten times on ten unique subsets. The statistical results are depicted in Table 5. The overall means of correct classification in the training (test) set, for

 © 2009 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

Table 4. Prediction performances for two LDA-based QSAR models (using non-stochastic and stochastic bond-type bilinear indices), in the training and test sets.

Matthews' corr. coeff. (C)	Accuracy "QTotal" (%)	Sensitivity "hit rate" (%)	Specificity (%)	False positive rate "false alarm rate" (%)	
Non-stochastic bond-type bilinear indices (Eq. 16)					
Training set	0.80		90.11	90.00	87.81	9.80
Predicting set	0.86		93.75	90.00	90.00	4.55
Stochastic bond-type linear bilinear indices (Eq. 17)					
Training set	0.76		87.91	87.50	85.37	11.77
Predicting set	0.71		87.50	80.00	80.00	9.09

Table 5. Results of the ten-*fold full* cross-validation procedure.

Groups	Q%[a]	λ	D^2	F	Q%[b]	Q%[a]	λ	D^2	F	Q%[b]
	Eq. 16 (non-stochastic indices)					Eq. 17 (stochastic indices)				
1	88.89	0.44	5.04	15.73	80.00	87.65	0.40	6.04	18.85	80.00
2	91.46	0.45	4.86	15.33	77.78	86.59	0.42	5.43	17.12	88.89
3	89.02	0.49	4.10	12.94	88.89	86.59	0.44	4.97	15.69	88.89
4	89.02	0.46	4.64	14.65	88.89	89.02	0.40	5.92	18.68	77.78
5	90.24	0.49	4.12	13.01	88.89	86.59	0.44	5.08	16.03	100.00
6	90.24	0.47	4.41	13.90	77.78	86.59	0.42	5.40	17.04	77.78
7	89.02	0.45	4.81	15.17	88.89	86.59	0.39	6.26	19.74	88.89
8	89.02	0.44	4.98	15.71	88.89	87.80	0.41	5.65	17.83	77.78
9	89.02	0.47	4.40	13.89	88.89	86.59	0.43	5.15	16.24	88.89
10	90.36	0.48	4.28	13.73	100.00	86.75	0.41	5.79	18.56	75.00
Mean	89.63	0.47	4.56	14.41	86.89	87.07	0.42	5.57	17.58	84.39
SD	0.88	0.02	0.35	1.06	6.74	0.83	0.02	0.43	1.37	7.91

[a] Globally good classification from both models in training (90% of the data) set.
[b] Globally good classification from both models in test (10% of the data) set.

Eqs. 16 and 17, were 89.63% (86.89%) and 87.07% (84.39%), correspondingly. The result for predictions on the 10% *full* cross-validation test evidenced the quality (robustness, stability, and predictive power) of the models.

3.1.2 Simulation of a Ligand-Based Virtual Screening Experiment for "Novel" Trichomonacidals Generation. First External Validation

One of the main features, which any theoretical approach to drug discovery needs, is the identification of active compounds from never-used databases of chemicals. This search can be understood as an alternative to screening approaches to drug discovery. By means of the procedures mentioned above, instead of assaying a large number of chemicals in a series of biological tests, one "virtually assays" these compounds by evaluating their activities with the models developed to this effect; this process is known today as computational (*virtual* or *in silico*) screening [142–144]. *Virtual* screening techniques may be classified according to their particular modeling of molecular recognition and to the type of algorithm used in database searching [69, 142, 143]. If the target (or at least its active site) 3-D structure is known, one of the structure-based *virtual* screening methods can be applied. In contrast, ligand-based methods are founded on the principle of similarity, namely, similar compounds are assumed to produce similar effects. Nevertheless, the absence of a receptor 3-D structure is the main reason for the application of ligand-based methods [68, 145, 146]. Because of the former fundamental facts, ligand-based *virtual* screening becomes our working philosophy.

In order to test the TOMOCOMD-CARDD potentialities, toward the ligand-based *virtual* screening of trichomonacidals, we have selected a series of 12 compounds whose activities against Tv have been already proved by several researchers [147, 148]. They all were evaluated ("screened") with models 16 and 17 as active/inactive ones. Their structures as well as the results of the classification are shown in Table 6.

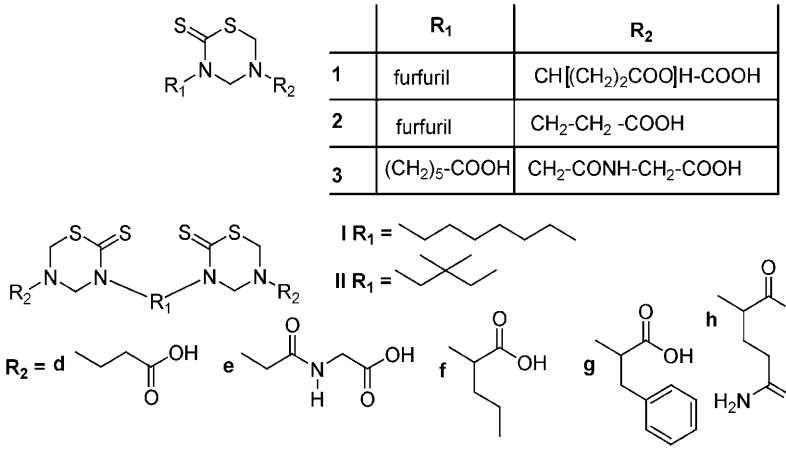
As observed, almost all the compounds were correctly classified by both models.

Equation 16 erroneously classified three (3) compounds as false negative, thus achieving a 75.00% (9/12) of correct classification, while Eq. 17 "let out" two (2) active compounds as false negative for yielding a 83.33% (10/12) of correct classification. The former model validation process is important, if we take into account that the predictive ability of a QSAR model can be estimated by using only an external test set of compounds that was not employed to build the model [132, 140, 141].

The next step in this approach would be the inclusion of these "novel" compounds in the training set and the development of a new discriminant model. This new model could be significantly different from the previous one, due to the inclusion of new structural patterns, but it also would be able to recognize a greater number of "strange compounds" as trichomonacidals. Therefore, the derivation of a classifier model is considered an iterative process, in which novel compounds with novel structural features are incorporated into the training set, for improving the quality of the so-developed models.

 © 2009 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

Table 6. Identification of chemicals extracted from literature as either active or inactive toward the antitrichomonal action, by using LDA-based QSAR models in a simulated ligand-based *virtual*-screening experiment.



	R_1	R_2
1	furfuril	$CH[(CH_2)_2COO]H$-COOH
2	furfuril	CH_2-CH_2 -COOH
3	$(CH_2)_5$-COOH	CH_2-CONH-CH_2-COOH

Comp.	Ref.[a]	$\Delta P\%$[b]	$\Delta P\%$[c]	Antitrichomonal activity (μg/mL) **100**	**10**	**1**
1	Ochoa *et al.*1999 [147]	76.57	79.43	100[d]	100[d]	(87)[e]
2		38.00	13.68	100[d]	94[d]	(59)[e]
3		12.46	**−68.36**	100[d]	100[d]	(65)[e]
Ie	Coro *et al.*, 2005 [148]	63.79	30.18	95.2[d]	(55.1)[e]	0
Ig		**−6.20**	4.78	100[d]	88.6[d]	(2.2)[e]
Ih		18.87	**−27.84**	100[d]	(46.8)[e]	0
Ij		75.14	85.15	92.7[d]	(30.2)[e]	0
Ik		31.58	37.42	100[d]	(50.2)[e]	(15.6)[e]
IId		−71.99	−89.10	Inactive		
IIe		46.06	66.45	100[d]	(84.0)[e]	(27.1)[e]
IIf		**−92.35**	10.95	100[d]	(64.4)[e]	0
IIg		**−30.82**	54.47	100[d]	(82.2)[e]	(38.8)[e]

The molecular structures of the compounds represented with bold figures are shown at the top of this table.
[a] Bibliographical references from where molecules together with its *in vitro* activities were taken.
[b] Antitrichomonal activity predicted by Eq 16: $\Delta P\% = [P(\text{active}) - P(\text{inactive})] \times 100$.
[c] Antitrichomonal activity predicted by Eq. 17: $\Delta P\% = [P(\text{active}) - P(\text{inactive})] \times 100$.
[d] Percentage of reduction of Tv or cytocidal activity at the indicated doses at 24 h.
[e] Specific activity against Tv (in parentheses) expressed as percentages of growth inhibition or cytostatic activity at 24 h.

3.2 QSAR Models as Virtual Shortcuts to New Antitrichomonal Compounds. From Dry Selection to Wet Evaluation

The massive cost involved in the development of new drugs, together with the low effectiveness of traditional assays in drug discovery, highlights the need for a "sea change" in the drug-discovery paradigm. Computational – *in silico* – screening of large databases, considering the use of predictive QSAR models, has emerged as an interesting alternative to High-Throughput Screening (HTS) and as an important drug-discovery tool [149, 150].

In order to test the reach of TOMOCOMD-CARDD and LDA strategies, for detecting new antiprotozoan compounds, we predicted the biological activity of eight (8) analogues of 4-[5-(dialkylamino)pentyl]quinoxalinones, which were provided by one of our synthetic research teams from IQM, CSIC, Spain [151]. The structures of these compounds are depicted in Figure 2.

All these compounds were initially screened with the QSAR models 16 and 17, then they were assayed *in vitro* in order to corroborate the predictions against Tv. Table 7 summarizes the theoretical and biological achievements.

As expected, a rather good agreement was observed between the theoretical predictions and the observed activity for both, active and inactive compounds. Our trained LDA-based QSAR models (Eqs. 16 and 17) were capable of successfully classifying seven out of eight compounds yielding (both) an accuracy of 87.50%.

As for the *in vitro* experiments, it is to be noted that almost all the compounds [VAM2-(3–8)] exhibited pronounced cytocidal activities of 100% at the concentration of 100 μg/mL and at 24 h (48 h), but VAM2–2: 99.37%

 © 2009 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

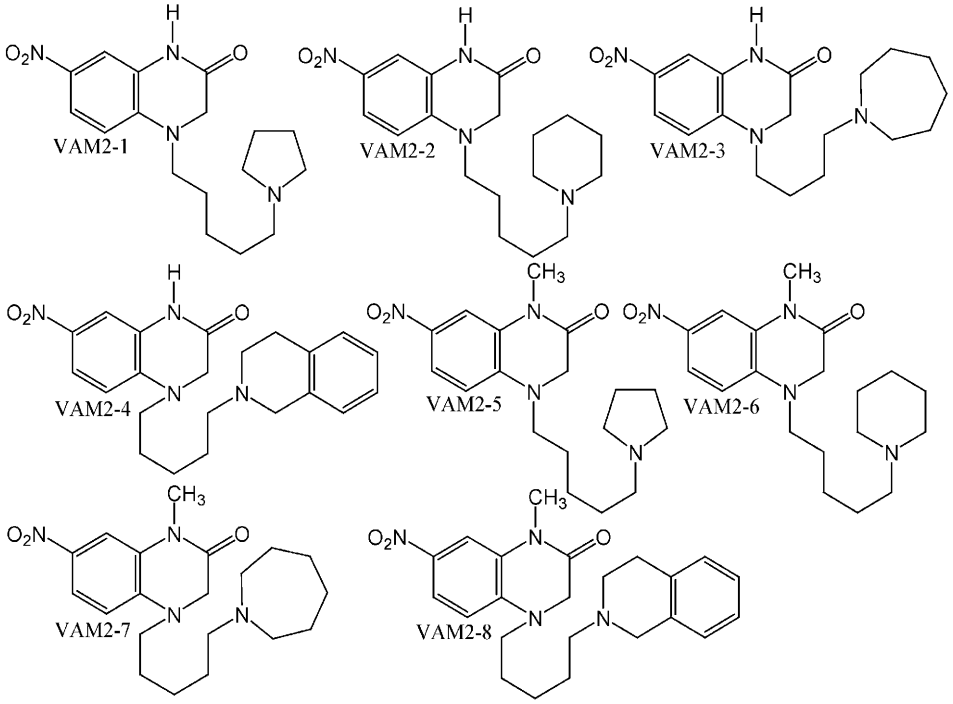


Figure 2. Structures of novel trichomonacidals discovery by ligand-based *virtual* screening LDA-assisted models.

Table 7. Results of the computational evaluation using LDA-assisted QSAR models and percentages of cytostatic and/or cytocidal activity (parenthesis) for the three concentrations assayed *in vitro* against Tv.

Compound[a]	Theoretical results					*In vitro* activity (μg/mL)[g]					
	Class[b]	$\Delta P\%$[c]	Class[d]	$\Delta P\%$[e]	Class[f]	$\%CA_{24\,h}$ ($\%C_{24\,h}$)			$\%CA_{48\,h}$ ($\%C_{48\,h}$)		
						100	**10**	**1**	**100**	**10**	**1**
VAM2-1	+	**71.66**	+	**81.90**	−	75.61	21.02	3.53	34.26	1.64	0
VAM2-2	+	68.19	+	81.04	+	**(99.37)**	20.94	0	**(100)**	5.74	0
VAM2-3	+	64.39	+	80.80	+	**(100)**	12.94	2.35	**(100)**	0	0
VAM2-4	+	79.65	+	93.28	+	**(100)**	83.76	3.53	**(100)**	44.06	0
VAM2-5	+	63.78	+	83.31	+	**(100)**	**(89.25)**	0	**(100)**	67.7	4.1
VAM2-6	+	59.58	+	82.52	+	**(100)**	**(92.63)**	0	**(100)**	86.52	0
VAM2-7	+	55.02	+	82.29	+	**(100)**	**(91.61)**	10.98	**(100)**	70.41	2.87
VAM2-8	+	73.66	+	93.83	+	**(100)**	70.98	4.71	**(100)**	23.28	4.1
MTZ	+	98.36	+	98.00	+	**(100)**	**(99.1)**	**(98.0)**	**(100)**	**(100)**	**(99.5)**

[a] The molecular structures of the compounds represented with codes are shown in Figure 2.
[b] *In silico* classification obtained from models Eq. 16 using non-stochastic bond-type bilinear indices.
[c] Results for the classification of compounds obtained from models Eq. 16: $\Delta P\% = [P(\text{active}) - P(\text{inactive})] \times 100$.
[d] *In silico* classification obtained from models Eq. 17 using stochastic bond-type bilinear indices.
[e] Results for the classification of compounds obtained from models Eq. 17: $\Delta P\% = [P(\text{active}) - P(\text{inactive})] \times 100$.
[f] Observed (experimental activity) classification against Tv.
[g] Pharmacological activity of each tested compound, which was added to the cultures at doses of 100, 10, and 1 μg/mL: $\%CA_{\#}$ = cytostatic activity $_{(24\text{ or }48\,h)}$ and ($\%C_{\#}$) = cytocidal activity $_{(24\text{ or }48\,h)}$. MTZ = Metronidazole (concentrations for MTZ were 2, 1, and 0.5 μg/mL, respectively).

(100%). It is remarkable that these compounds did not show toxic activity in macrophage cultivations at this concentration. Moreover, as observed in Table 7, compounds [VAM2-(5–7)] maintained a high trichomonacidal activity (89.25, 92.63, and 91.61%, respectively) as well as non-specific cytotoxicity at concentrations of 10 μg/mL but only at 24 h, while compounds [VAM2-(2–4)] and VAM2–8 showed non-selective trichomonacidal activity, but low-to-moderate cytostatic activity at both periods.

Although compounds [VAM2-(5–7)] were active at higher doses than MTZ (reference drug), this result leaves a door opened to the *virtual* variational study of their structures, in order to improve their antitrichomonal activity.

 © 2009 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

4 Concluding Remarks

Once more, QSAR studies arise as an efficient alternative to time-consuming and costly conventional HTS, *in vitro* or/and *in vivo* techniques. Particularly, there were demonstrated the possibilities of TOMOCOMD-CARDD's MDs and LDA, for generating simple and robust models, capable not only of identifying well-known and established trichomonacidals, but also of discovering new candidates. The latter is important, not only because it was possible to describe such a complex biological phenomenon by a few variables, but also by the attempt to provide, at least, a partial solution to the generation of adequate drugs, competing with MTZ in the treatment of trichomoniasis, so needed nowadays.

Acknowledgement

Yovani Marrero-Ponce (M.-P. Y.) acknowledges the Valencia University for kind hospitality during the second semester of 2007 M.-P. Y. thanks are given to the Generalitat Valenciana (Spain) for partial financial support as well as the program "Estades Temporals per an Investigadors Convidats" for a fellowship to work at Valencia University (2006 – 2007). Some authors' thanks support from Spanish MEC (Project Reference: SAF2006 – 04698). Finally, F. T. thanks support from Spanish MEC (Project No. CTQ2004-07768-C02-01/BQU and CCT005-07-00365) and. EU (Program Feder). We are also indebted to the journal Editor Professor Dr. Gisbert Schneider for his kind attention. Finally, but not less, M-P, Y thanks are given to the projects entitle "Strengthening postgraduate education and research in Pharmaceutical Sciences". This project is funded by the Flemish Interuniversity Council (VLIR) of Belgium.

References

[1] D. Petrin, K. Delgaty, R. Bhatt, G. Garber, *Clin. Microbiol. Rev.* **1998**, *11*, 300 – 317.
[2] W. J. Cates, *Sex. Transm. Dis.* **1999**, *26*, S2 – S7.
[3] World-Health-Organization, Global program on AIDS, An overview of selected curable sexually transmitted diseases, World Health Organization, Geneva, Switzerland, **1995**.
[4] M. T. Brown, *Practitioner* **1972**, *209*, 639 – 644.
[5] R. D. Catterall, *Med. Clin. North Am.* **1972**, *56*, 1203 – 1209.
[6] A. R. Wisdom,E. M. C. Dunlop, *Br. J. Vener. Dis.* **1965**, *41*, 90 – 96.
[7] S. García,D. A. Bruckner, *Diagnostic Medical Parasitology*, American Society for Microbiology, Washington (DC) **1993**, pp. 84 – 91.
[8] M. F. Rein, *Trichomoniasis*, R. Goldsmith, D. Heyneman (Eds.), Santafé de Bogotá **1995**.
[9] I. T. Gram, M. Macaluso, J. Churchill, H. Stalsberg, *Cancer Causes Control* **1992**, *3*, 231 – 236.
[10] Z. F. Zhang,C. B. Begg, *Int. J. Epidemiol.* **1994**, *23*, 682 – 690.
[11] M. Viikki, E. Pukkala, P. Nieminen,M. Hakama, *Acta Oncol.* **2000**, *39*, 71 – 75.
[12] B. M. Kharsany, A. A. Hoosen, J. Moodley, J. Bagaratee, E. Gouws, *Genitourin. Med.* **1993**, *69*, 357 – 360.
[13] W. Cates, M. R. Joesoef, M. B. Goldman, *Am. J. Obstet. Gynecol.* **1993**, *169*, 341 – 346.
[14] F. Grodstein, M. B. Goldman, D. W. Cramer, *Am. J. Epidemiol.* **1993**, *137*, 577 – 584.
[15] D. E. Soper, R. C. Bump,W. G. Hurt, *Am. J. Obstet. Gynecol.* **1990**, *163*, 1016 – 1023.
[16] M. F. Cotch, Program and abstracts of the 30th Interscience Conference on Antimicrobial Agents and Chemotherapy, Abs. 681, Vaginal infections and prematurity study group. Carriage of *Trichomonas vaginalis* (Tv) is associated with adverse pregnancy outcome, Washington DC **1990**.
[17] H. Minkoff, A. N. Grunebaum, R. H. Schwarz, J. Feldman, M. Cummings, W. Crombleholme, L. Clark, G. Pringle, W. M. McCormack, *Am. J. Obstet. Gynecol.* **1984**, *150*, 965 – 972.
[18] K. B. Fowler, R. F. Pass, *J. Infect. Dis.* **1991**, *164*, 259 – 264.
[19] M. Laga, A. Manoka, M. Kivuvu, B. Malele, M. Tuliza, N. Nzila, J. Goeman, F. Behets, V. Batter, M. Alary, W. L. Heyward, R. W. Ryder, P. Piot, *AIDS* **1993**, *7*, 95 – 102.
[20] F. Sorvillo, P. Kerndt, *Lancet.* **1998**, *351*, 213 – 214.
[21] B. M. Honigberg, V. M. King, *J. Parasitol.* **1964**, *50*, 345 – 364.
[22] M. Müller, *Symp. Soc. Gen. Microbiol.* **1980**, *30*, 127 – 142.
[23] M. Müller, *Acta Univ. Carol. Biol.* **1987**, *30*, 249 – 260.
[24] J. G. Lossick, H. L. Kent, *Am. J. Obstet. Gynecol.* **1991**, *165*, 1217 – 1222.
[25] C. Cosar, L. Julou, *Ann. Inst. Pasteur* **1959**, *96*, 238 – 241.
[26] N. Yarlett, N. C. Yarlett, D. Lloyd, *Biochem. Pharmacol.* **1986**, *35*, 1703 – 1708.
[27] J. H. Tocher, D. I. Edwards, *Biochem. Pharmacol.* **1994**, *48*, 1089 – 1094.
[28] M. H. Nielsen, *Acta Pathol. Microbiol. Scand B* **1976**, *84*, 93 – 100.
[29] Centers for Disease Control and Prevention, Sexually transmitted diseases treatment guidelines, *Morb. Mortal. Wkly. Rep.* **1993**, *42* 70.
[30] A. Garcia-Leverde, L. de Bonila, *Am. J. Trop. Med. Hyg.* **1975**, *24*, 781 – 783.
[31] S. J. Powell, L. Macleod, A. J. Wilmot, R. Elsdon-Dew, *Lancet.* **1966**, *ii*, 1329 – 1331.
[32] J. Scheider, *Bull. Soc. Pathol. Exot.* **1961**, *54*, 84 – 93.
[33] S. M. Townson, P. F. L. Boreham, P. Upcroft, J. A. Upcroft, *Acta Trop.* **1994**, *56*, 173 – 194.
[34] R. Knight, *J. Antimicrob. Chemother.* **1980**, *6*, 577 – 593.
[35] J. Kulda, M. Vojtěchovská, J. Tachezy, P. Demeš, E. Kunzová, *Br. J. Vener. Dis.* **1982**, *58*, 394 – 399.
[36] J. G. Lossick, M. Müller, T. E. Gorrell, *J. Infect. Dis.* **1986**, *153*, 948 – 955.
[37] J. G. Meingassner, J. Thurner, *Antimicrob. Agents Chemother.* **1979**, *15*, 254 – 257.
[38] T. Meri, T. S. Jokiranta, L. Suhonen, S. Meri, *J. Clin. Microbiol.* **2000**, *38*, 763 – 767.
[39] M. Müller, J. G. Meingassner, W. A. Miller,W. J. Ledger, *Am. J. Obstet. Gynecol.* **1980**, *138*, 808 – 812.
[40] D. M. Brown, J. A. Upcroft, H. N. Dodd, N. Chen, P. Upcroft, *Mol. Biochem. Parasitol.* **1999**, *98*, 203 – 214.
[41] J. Kulda, J. Čerkasov, P. Demeš, A. Čerkasovová, *Exp. Parasitol.* **1984**, *57*, 93 – 103.

 © 2009 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

[42] J. Kulda, J. Tachezy, A. Čerkasovová, *J. Eukaryot. Microbiol.* **1993**, *40*, 262 – 269.
[43] $\bar{u} = \sum_{i=1}^{m} u^i \bar{e}_i$ J. Tachezy, J. Kulda, E. Tomková, *Parasitology* **1993**, *106*, 31 – 37.
[44] I. de Carneri, G. Achilli, G. Monti, F. Trane, *Lancet* **1969**, *2*, 1308 – 1309.
[45] J. G. Meingassner, H. Mieth, R. Czok, D. G. Lindmark, M. Müller, *Antimicrob. Agents Chemother.* **1978**, *13*, 1 – 3.
[46] J. Kulda, *Int. J. Parasitol*. **1999**, *29*, 199 – 212.
[47] H. Gillette, G. P. Schmid, D. Moswe, XIIIth Meeting of the International Society of Sexually Transmitted Disease Research, Metronidazole-resistant Trichomonas vaginalis, a case series, Denver, July 11 – 14 **1999**.
[48] J. G. Lossick, H. L. Kent, *Am. J. Obstet. Gynecol.* **1991**, *165*, 1217 – 1222.
[49] B. C.-O. C. Group, *Int. J. STD AIDS* **1992**, *3*, 24 – 27.
[50] M. Dan, J. D. Sobel, *Infect. Dis. Obstet. Gynecol.* **1996**, *4*, 77 – 84.
[51] I. H. Ahmed-Jushuf, A. E. Murray, J. McKeown, *Genitourin. Med.* **1988**, *64*, 25 – 29.
[52] J. H. I. Grossman, R. P. Galask, *Obstet. Gynecol.* **1990**, *76*, 521 – 522.
[53] D. A. Lewis, L. Habgood, R. White, K. F. Barker, S. M. Murphy, *Int. J. STD AIDS* **1997**, *8*, 780 – 784.
[54] J. G. Lossick, H. L. Kent, *Am. J. Obstet. Gynecol.* **1991**, *165*, 1217 – 1222.
[55] J. G. Lossick, M. Muller, T. E. Gorrell, *J. Infect. Dis.* **1986**, *153*, 948 – 955.
[56] M. P. Dombrowski, R. J. Sokol, W. J. Brown, R. A. Bronsteen, *Obstet. Gynecol.* **1987**, *69*, 524 – 525.
[57] G. Saurina, W. M. McCormack, D. Landman, XIIIth Meeting of the International Society of Sexually Transmitted Disease Research, A study of the prevalence of resistant trichomoniasis and response to treatment in Brooklyn, NY Denver, July 11 – 14 **1999**.
[58] M. Dan, J. D. Sobel, *Infect. Dis. Obstet. Gynecol.* **1996**, *4*, 77 – 84.
[59] E. M. Narcisi, W. E. Secor, *Antimicrob. Agents Chemother.* **1996**, *40*, 1121 – 1125.
[60] E. T. Houang, Z. Ahmet, A. G. Lawrence, *Sex. Transm. Dis.* **1997**, *24*, 116 – 119.
[61] R. S. Pattman, M. S. Sprott, A. M. Kerns, M. Earnshaw, *Genitourin. Med.* **1989**, *65*, 274 – 275.
[62] C. A. Wong, P. D. Wilson, T. A. Chew, *Aust. NZ J. Obstet. Gynaecol.* **1990**, *30*, 169 – 171.
[63] C. H. I. Livengood, J. G. Lossick, *Obstet. Gynecol.* **1991**, *78*, 954 – 956.
[64] P. G. Watson, R. S. Pattman, *Int. J. STD AIDS* **1996**, *7*, 296 – 297.
[65] P. Nyirjesy, J. D. Sobel, M. V. Weitz, *Clin. Infect. Dis.* **1998**, *26*, 986 – 988.
[66] P. Nyirjesy, M. V. Weitz, S. P. Gelone,T. Fekete, *Lancet* **1995**, *346*, 1110.
[67] E. Estrada, A. Peña, *Bioorg. Med. Chem.* **2000**, *8*, 2755 – 2770.
[68] E. Estrada, E. Uriarte, A. Montero, M. Teijeira, L. Santana, E. De Clercq, *J. Med. Chem.* **2000**, *43*, 1975 – 1985.
[69] R. K. Scott, Informatics integration: The bedrock of NCE selection. *Biosilico* **2003**, *1*, 14 – 17.
[70] Y. Marrero-Ponce,V. Romero, TOMOCOMD (TOpological MOlecular COMputer Design) for Windows, 1.0, 2002, TOMOCOMD software, Central University of Las Villas., version 1.0 is a preliminary experimental version; in future a professional version will be obtained upon request to Yovani Marrero Ponce, yovanimp@qf.uclv.edu.cu or ymarrero77@yahoo.es (for more details see www.uv.es/yoma).
[71] Y. Marrero-Ponce, *Molecules* **2003**, *8*, 687 – 726.
[72] Y. Marrero-Ponce, *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 2010 – 26.
[73] Y. Marrero-Ponce, *Bioorg. Med. Chem.* **2004**, *12*, 6351 – 6369.
[74] Y. Marrero-Ponce, J. A. Castillo-Garit, F. Torrens, V. Romero-Zaldivar, E. Castro, *Molecules* **2004**, *9*, 1100 – 1123.
[75] Y. Marrero-Ponce, H. G. Díaz, V. Romero, F. Torrens, E. A. Castro, *Bioorg. Med. Chem.* **2004**, *12*, 5331 – 5342.
[76] J. A. Castillo-Garit, Y. Marrero-Ponce, F. Torrens, R. Rotondo, *J. Mol. Graphics. Model.* 2006. *26*, 32 – 47.
[77] Y. Marrero-Ponce, J. A. Castillo-Garit, *J. Comput. Aid. Mol. Des.* **2005**, *19*, 369 – 383.
[78] Y. Marrero-Ponce, M. A. Cabrera, V. Romero, E. Ofori, L. A. Montero, *Int. J. Mol. Sci.* **2003**, *4*, 512 – 36.
[79] Y. Marrero-Ponce, M. A. Cabrera, V. Romero, D. H. González, F. Torrens, *J. Pharm. Pharmaceut. Sci.* **2004**, *7*, 186 – 199.
[80] Y. Marrero-Ponce, M. A. Cabrera, V. Romero-Zaldivar, M. Bermejo, D. Siverio, F. Torrens, *Internet Electron. J. Mol. Des.* **2005**, *4* 124 – 150.
[81] Y. Marrero-Ponce, J. A. Castillo-Garit, E. Olazabal, H. S. Serrano, A. Morales, N. Castanedo, F. Ibarra-Velarde, A. Huesca-Guillen, A. M. Sanchez, F. Torrens, E. A. Castro, *Bioorg. Med. Chem*. **2005**, *13*, 1005 – 1020.
[82] Y. Marrero-Ponce, J. A. Castillo-Garit, E. Olazabal, H. S. Serrano, A. Morales, N. Castanedo, F. Ibarra-Velarde, A. Huesca-Guillen, E. Jorge, A. del Valle, F. Torrens,E. A. Castro, *J. Comput. Aid. Mol. Des.* **2004**, *18*, 615 – 34.
[83] Y. Marrero-Ponce, A. Huesca-Guillen, F. Ibarra-Velarde, *J. Mol. Struct. (Theochem.)* **2005**, *717*, 67 – 79.
[84] Y. Marrero-Ponce, A. Montero-Torres, C. R. Zaldivar, M. I. Veitia, M. M. Perez, R. N. Sanchez, *Bioorg. Med. Chem.* **2005**, *13*, 1293 – 1304.
[85] Y. Marrero-Ponce, R. Medina-Marrero, F. Torrens, Y. Martinez, V. Romero-Zaldivar, E. A. Castro, *Bioorg. Med. Chem.* **2005**, *13*, 2881 – 99.
[86] Y. Marrero-Ponce, R. Medina-Marrero, Y. Martinez, F. Torrens, V. Romero-Zaldivar, E. A. Castro, *J. Mol. Model.* **2006**, *12*, 255 – 271.
[87] Y. Marrero-Ponce, D. Nodarse, H. D. González, R. Ramos de Armas, V. Romero-Zaldivar, F. Torrens, E. Castro, *Int. J. Mol. Sci.* **2004**, *5*, 276 – 293.
[88] Y. Marrero-Ponce, J. A. Castillo-Garit, D. Nodarse, *Bioorg. Med. Chem.* **2005**, *13*, 3397.
[89] Y. Marrero-Ponce, R. Medina, E. A. Castro, R. de Armas, H. González, V. Romero, F. Torrens, *Molecules* **2004**, *9*, 1124 – 1147.
[90] Y. Marrero-Ponce, R. Medina-Marrero, J. A. Castillo-Garit, V. Romero-Zaldivar, F. Torrens, E. A. Castro, *Bioorg. Med. Chem.* **2005**, *13*, 3003 – 3015.
[91] Y. Marrero-Ponce, F. Torrens, Y. J. Alvarado, R. Rotondo,V. Romero-Zaldivar, *J. Comp.-Aid. Mol. Des.* **2006**, *20*, 685 – 701.
[92] Y. Marrero-Ponce, M. T. H. Khan, G. M. Casañola-Martin, A. Ather, K. M. Khan, F. Torrens, R. Rotondo, *J. Comput. Aid. Mol. Design.* **2007**, *21*, 167 – 188.
[93] G. M. Casañola-Martin, M. T. H. Khan, Y. Marrero-Ponce, A. Ather, S. Sultan, F. Torrens, R. Rotondo, *Bioorg. Med. Chem.* **2007**, *15*, 1483 – 1503.
[94] D. H. Rouvray (Ed.), *Chemical Applications of Graph Theory*, Academic Press, London **1976**, pp. 180 – 181.

 © 2009 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

[95] N. Trinajstić (Ed.), *Chemical Graph Theory*, CRC Press, Boca Raton, FL **1992**.
[96] E. Estrada, *J. Chem. Inf. Comput. Sci*. **1995**, *35*, 31 – 33.
[97] E. Estrada, A. Ramirez, *J. Chem. Inf. Comput. Sci*. **1996**, *36*, 837 – 43.
[98] E. Estrada, *J. Chem. Inf. Comput. Sci*. **1996**, *36*, 844 – 49.
[99] E. Estrada, N. Guevara, I. Gutman, *J. Chem. Inf. Comput. Sci*. **1998**, *38*, 428 – 31.
[100] E. Estrada, *J. Chem. Inf. Comput. Sci*. **1999**, *39*, 1042 – 48.
[101] E. Estrada, E. Molina, *J. Mol. Graph. Model.* **2001**, *20*, 54 – 64.
[102] R. Todeschini, V. Consonni, *Handbook of Molecular Descriptors*, Weinheim, Wiley-VCH, Germany **2000**, pp. 668.
[103] C. H. Edwards, D. E. Penney, *Elementary Linear Algebra*, Prentice-Hall, Englewood Cliffs, New Jersey, USA **1988**.
[104] Y. Marrero Ponce, *J. Chem. Inf. Comput. Sci*. **2004**, *44*, 2010 – 2026.
[105] E. Estrada, S. Vilar, E. Uriarte,Y. Gutierrez, *J. Chem. Inf. Comput. Sci*. **2002**, *42*, 1194 – 203.
[106] E. Estrada, A. Peña, R. Garcia-Domenech, *J. Comput. Aid. Mol. Des.* **1998**, *12*, 583 – 595.
[107] V. M. Potapov, *Stereochemistry*, Mir, Moscow **1978**.
[108] R. Wang, Y. Gao, L. Lai, *Perspect. Drug Disc. Des.* **2000**, *19*, 47 – 66.
[109] P. Ertl, B. Rohde, P. Selzer, *J. Med. Chem.* **2000**, *43*, 3714 – 7.
[110] A. K. Ghose, G. M. Crippen, *J. Chem. Inf. Comput. Sci*. **1987**, *27*, 21 – 35.
[111] K. J. Miller, *J. Am. Chem. Soc.* **1990**, *112*, 8533 – 8542.
[112] J. Gasteiger,M. Marsili, *Tetrahedron Lett.* **1978**, *19*, 3181 – 3184.
[113] L. Pauling, *The Nature of Chemical Bond*, Cornell University Press, Ithaca (New York) **1939**, pp. 2 – 60.
[114] L. B. Kier, L. H. Hall, *Molecular Connectivity in Structure – Activity Analysis*, Research Studies Press, Letchworth, UK **1986**, pp. 262.
[115] P. G. Hoel, *Introducción a la Estadistica Matemática*, Editorial Pueblo y Educación, La Habana **1978**, pp. 18 – 22.
[116] K. F. Riley, M. P. Hobson, S. J. Vence, *Mathematical Methods for Physics and Engineering*, Cambridge University Press, Cambridge, **1998**, pp. 228 – 236.
[117] E. Hernández, *Álgebra y Geometría*, Universidad Autonoma de Madrid, Madrid **1987**, pp. 521 – 544.
[118] J. de Burgos-Román, *Álgebra y Geometría Cartesíana*, McGraw-Hill Interamericana de España, España **2000**, pp. 208 – 246.
[119] J. de Burgos-Román, *Curso de Álgebra y Geometría*, Alambra Longman, Ed., Madrid **1994**, pp. 638 – 684.
[120] G. Werner, *Linear Algebra*, Springer-Verlag, New York **1981**, pp. 261 – 288.
[121] P. D. Walker, P. G. Mezey, *J. Am. Chem. Soc.* **1993**, *115*, 12423.
[122] D. J. Klein, *Internet Electron. J. Mol. Des.* **2003**, *2*, 814 – 834.
[123] J. H. Van Vleck, A. Sherman, *Rev. Mod. Phys.* **1935**, *7*, 167 – 228.
[124] R. Todeschini, V. Consonni, M. Pavan, Dragon Software, *version 2.1*. **2002**.
[125] R. Todeschini, P. Gramatica, *Perspect. Drug Disc. Des.* **1998**, *9 – 11*, 355 – 380.
[126] V. Consonni, R. Todeschini, M. Pavan, *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 682 – 92.
[127] M. Negwer, *Organic-Chemical Drugs and their Synonyms*, Akademie-Verlag, Berlin **1987**.
[128] S. Budavari, M. O'Neil, Ann Smith, P. Heckelman, J. Obenchain. *The Merck Index on CD-ROM*, Chapman and Hall and Merck & Co., Inc. **1999.**
[129] H. van de Waterbeemd, (Ed.), *Chemometric Methods in Molecular Des.*, VCH Publishers, Weinheim **1995**, pp. 265 – 288.
[130] STATISTICA (data analysis software system) vs 6.0, StatSoft, Inc. www.statsoft.com. 2001.
[131] E. Estrada, G. Patlewicz, *Croat. Chim. Acta* **2004**, *77*, 203 – 211.
[132] S. Wold, L. Erikson, in: *Chemometric Methods in Molecular Des.*, H. van de Waterbeemd (Ed.), VCH Publishers, New York **1995**, pp. 309 – 318.
[133] P. Baldi, S. Brunak, Y. Chauvin, C. A. Andersen, H. Nielsen, Assessing the Accuracy of Prediction Algorithms for Classification: an Overview. *Bioinformatics* **2000**, *16*, 412 – 424.
[134] V. V. Kouznetsov, C. J. Rivero, P. C. Ochoa, E. Stashenko, J. R. Martínez, P. D. Montero, R. J. J. Nogal, P. C. Fernández, S. S. Muelas, B. A. Gómez, A. Bahsas, L. Amaro, *J. Arch. Pharm.* **2005**, *1*, 338.
[135] V. V. Kouznetsov, M. L. Y. Vargas, B. Tibaduiza, C. Ochoa, P. D. Montero, R. J. J. Nogal, C. Fernández, S. Muelas, A. Gómez, A. Bahsas, J. Amaro-Luis, *J. Arch. Pharm.* **2004**, *337*, 127 – 132.
[136] J. Gálvez, R. Garcia-Domenech, J. V. de Julián-Ortiz, R. Soler, *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 272 – 284.
[137] R. A. Cercos-del-Pozo, F. Pérez-Giménez, M. T. Salabert-Salvador, F. J. Garcia-March, *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 178 – 184.
[138] J. Gálvez, R. García, M. T. Salabert, R. Soler, *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 520 – 525.
[139] R. A. Johnson, D. W. Wichern, *Applied Multivariate Statistical Analysis*, Prentice-Hall, New Jersey **1988**.
[140] A. Golbraikh,A. Tropsha, *J. Mol. Graph. Model.* **2002**, *20*, 269 – 276.
[141] K. Rose, L. H. Hall, L. B. Kier, *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 651 – 666.
[142] J. Xu, A. Hagler, *Molecules* **2002**, *7*, 566 – 700.
[143] M. H. J. Seifert, K. Wolf, D. Vitt, Virtual High-Throughput *in silico* Screening. *Biosilico* **2003**, *1*, 143 – 149.
[144] C. Watson, Predictive *in silico* Models in Drug Discovery. *Biosilico* **2003**, *1*, 83 – 84.
[145] J. W. Mc Farland, D. J. Gans, (Ed.), *Chemometric Methods in Molecular Des.*, VCH Publishers, New York **1995**, pp. 295 – 307.
[146] E. Estrada, E. Uriarte, *Curr. Med. Chem.* **2001**, *8*, 1573 – 1588.
[147] A. Ochoa, E. Pérez, R. Pérez, M. Suárez, E. Ochoa, H. Rodríguez, A. Gómez, S. Muelas, R. J. J. Nogal, R. A. Martínez, *Arzneim. Forsch./*Drug Res. **1999**, *49*, 764 – 769.
[148] J. Coro, R. Pérez, H. Rodríguez, M. Suárez, M. C. Vega, M. Rolón, D. Montero, J. J. Nogal, A. Gómez, *Bioorg. Med. Chem.* **2005**, *13*, 3413 – 3421.
[149] M. Lajiness, Molecular similarity-based methods for selecting compounds for screening in: D. H.Rouvray (Ed.), *Computacional Chemical Graph Theory*, Nova Science, New York **1990**, pp. 299 – 316.
[150] W. P. Walters, M. T. Stahl, M. A. Murcko, *Drug Discov. Today* **1998**, *3*, 160 – 178.
[151] Y. Marrero-Ponce, M. A. Martins Alho, V. Aran, J. A. Escario, R. Gracía Sánchez, J. J. Nogal-Ruiz, A. Meneses-Marcel, N. Rivera, *J. Med. Chem*., in progress.

 © 2009 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim