Ligand-Based Virtual Screening and in Silico Design of New Antimalarial Compounds Using Nonstochastic and Stochastic Total and Atom-Type Quadratic Maps

Yovani Marrero-Ponce,*,† Maité Iyarreta-Veitía,‡ Alina Montero-Torres,† Carlos Romero-Zaldivar,† Carlos A. Brandt,§ Priscilla E. Ávila,|| Karin Kirchgatter,|| and Yanetsy Machado†

Department of Pharmacy, Faculty of Chemical Pharmacy and Department of Drug Design, Chemical Bioactive Center, Central University of Las Villas, Santa Clara, 54830 Villa Clara, Cuba, Centre d'Etudes Pharmaceutiques, CNRS Biocis UMR 8076, Laboratoire de Synthese de Composes d'Interet Biologique, Faculté de Pharmacie, Université Paris-Sud 5, rue J.B. Clément, 92296 Châtenay-Malabry Cedex, France, Department of Organic Chemistry, Butantan Institute, Av. Vital Brasil 1500, Butantã, São Paulo, SP 05503-900, Brazil, and Núcleo de Estudos em Malária, Superintendência de Controle de Endemias (SUCEN), Av. Dr. Eneas de Carvalho Aguiar 470, Cerqueira César, São Paulo, SP 05403-000, Brazil

Received March 15, 2005

Malaria has been one of the most significant public health problems for centuries. It affects many tropical and subtropical regions of the world. The increasing resistance of *Plasmodium* spp. to existing therapies has heightened alarms about malaria in the international health community. Nowadays, there is a pressing need for identifying and developing new drug-based antimalarial therapies. In an effort to overcome this problem, the main purpose of this study is to develop simple linear discriminant-based quantitative structure−activity relationship (QSAR) models for the classification and prediction of antimalarial activity using some of the TOMOCOMD−CARDD (***TO***pological ***MO***lecular ***COM***puter ***D***esign−***C***omputer ***A***ided "***R***ational" ***D***rug ***D***esign) fingerprints, so as to enable computational screening from virtual combinatorial datasets. In this sense, a database of 1562 organic chemicals having great structural variability, 597 of them antimalarial agents and 965 compounds having other clinical uses, was analyzed and presented as a helpful tool, not only for theoretical chemists but also for other researchers in this area. This series of compounds was processed by a *k*-means cluster analysis in order to design training and predicting sets. Afterward, two linear classification functions were derived in order to discriminate between antimalarial and nonantimalarial compounds. The models (including nonstochastic and stochastic indices) correctly classify more than 93% of the compound set, in both training and external prediction datasets. They showed high Matthews' correlation coefficients, 0.889 and 0.866 for the training set and 0.855 and 0.857 for the test one. The models' predictivity was also assessed and validated by the random removal of 10% of the compounds to form a new test set, for which predictions were made using the models. The overall means of the correct classification for this process (leave group 10% full-out cross validation) using the equations with nonstochastic and stochastic atom-based quadratic fingerprints were 93.93% and 92.77%, respectively. The quadratic maps-based TOMOCOMD−CARDD approach implemented in this work was successfully compared with four of the most useful models for antimalarials selection reported to date. The developed models were then used in a simulation of a virtual search for Ras FTase (FTase = farnesyltransferase) inhibitors with antimalarial activity; 70% and 100% of the 10 inhibitors used in this virtual search were correctly classified, showing the ability of the models to identify new lead antimalarials. Finally, these two QSAR models were used in the identification of previously unknown antimalarials. In this sense, three synthetic intermediaries of quinolinic compounds were evaluated as active/inactive ones using the developed models. The synthesis and biological evaluation of these chemicals against two malaria strains, using chloroquine as a reference, was performed. An accuracy of 100% with the theoretical predictions was observed. Compound **3** showed antimalarial activity, being the first report of an arylaminomethylenemalonate having such behavior. This result opens a door to a virtual study considering a higher variability of the structural core already evaluated, as well as of other chemicals not included in this study. We conclude that the approach described here seems to be a promising QSAR tool for the molecular discovery of novel classes of antimalarial drugs, which may meet the dual challenges posed by drug-resistant parasites and the rapid progression of malaria illnesses.

1. BACKGROUND

Malaria remains one of the most serious health threats in the world, affecting 300−400 million people and claiming ca. 3 million lives each year.[1,2] As a result of the increasing prevalence of multidrug resistance of malaria parasites to standard chemotherapy, the discovery and use of nontraditional antimalarials with novel action mechanisms is becoming widespread.[3−5] Knowing the complexity and cost of the process of drug discovery, the use of "rational" search methodologies is recommended. Consequently, medicinal chemists are called to develop more efficient strategies for the search of novel candidates to be assayed as antimalarial

* To whom correspondence should be addressed. Fax: 53-42-281130, 281455. Tel.: 53-42-281192, 281473. E-mail: yovanimp@qf.uclv.edu.cu or ymarrero77@yahoo.es.
† Central University of Las Villas.
‡ Université Paris-Sud 5.
§ Butantan Institute.
|| Superintendência de Controle de Endemias.

10.1021/ci050085t CCC: $30.25
Published on Web 06/09/2005

drugs. In this sense, the computer-aided drug design approach emerges as a promising solution to this problem.[6−9] One of the major goals of such a design strategy is the identification, from large databases or libraries, of structural subsystems responsible for a specific biological activity. When computational approaches based on discrimination functions are used, it is possible to classify active compounds from inactive ones and to predict the biological activity of new lead compounds.[10−14] That is to say, cheminformatical in silico methods appear to be particularly rewarding in terms of both cost and time benefits and are easily integrated into the modern drug discovery process.[10−14]

To reduce costs, pharmaceutical companies have had to find new technologies to replace the old "hand-crafted" synthesis and testing of new chemical entities. Since 1980, with the advent of high throughput screening (HTS), automated techniques have made possible robotized screening. Through this process, hundreds of thousands of individual compounds can be screened per drug target per year. In conjunction with HTS technology, virtual (in silico) screening has become a main tool for identifying leads.[10−14] Virtual screening is actually one of the computational tools used to filter out unwanted chemicals from physical and in silico libraries.[10−14] Virtual screening techniques may be classified according to their particular modeling of molecular recognition and the type of algorithm used in database searching.[10,11] If the target 3D structure is known (or at least that of its active site), one of the structure-based virtual screening methods can be applied. These methods are based on the principle of complementarity: the receptor of a biologically active compound is complementary to the compound itself (i.e., a lock-and-key model).[10,11] By contrast, ligand-based methods are related in the principle of similarity; that is, similar compounds are assumed to produce similar effects. In this case, if one or more active chemicals are known, it is possible to search a database for similar but more potent molecules. This in silico approach will be used here in order to find predictive models for ligand-based virtual screening and rational design of antimalarial compounds.

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 (QSAR = quantitative structure−activity relationship; QSPR = quantitative structure property relationship), known as TOMOCOMD−CARDD (acronym of ***TO***pological ***MO***lecular ***COM***puter ***D***esign−***C***omputer-***A***ided "***R***ational" ***D***rug ***D***esign).[15]

This method has been developed to generate molecular fingerprints on the basis of the application of discrete mathematics and linear algebra theory to chemistry. In this sense, atom, atom-type, and total quadratic and linear molecular fingerprints have been defined in analogy to the quadratic and linear mathematical maps.[16,17] This (in silico) method has been successfully applied to predict several physical, physicochemical, and chemical properties of organic compounds.[16−19] In addition, TOMOCOMD−CARDD has been extended to considering 3D features of small-/medium-sized molecules on the basis of the trigonometric 3D-chirality correction factor approach.[20]

Later, the description of the significance interpretation and the comparison to other molecular descriptors was performed.[17,18] This approach describes changes over time in the electronic distribution throughout the molecular backbone. Specifically, the features of the *k*th total and local quadratic and linear indices were illustrated by examples of various types of molecular structures, including chain length and branching as well as the content of heteroatoms, and multiple bonds.[17,18] Additionally, the linear independence of the atom-type quadratic and linear fingerprints to 229 other 0D−3D molecular descriptors was demonstrated. In this sense, it was concluded that local TOMOCOMD−CARDD fingerprints are independent indices that contain important structural information to be used in QSPR/QSAR and drug design studies.[17,18]

The prediction of the pharmacokinetic properties of organic compounds is a problem that can also be addressed using this approach. In this sense, this method has been used to estimate the intestinal-epithelial transport of drugs in human adenocarcinoma of colon cell line type 2 (Caco-2) cultures for a heterogeneous series of drug-like compounds.[21−23] The obtained results suggested that the TOMOCOMD−CARDD descriptors were able to predict permeability values, and it proved to be a good tool for studying the oral absorption of drug candidates during the drug development process.

The TOMOCOMD−CARDD strategy has also been used for the selection of novel subsystems of compounds having a desired property/activity. In this sense, it was successfully applied to the virtual (computational) screening of novel anthelmintic compounds, which were then synthesized and evaluated in vivo on *Fasciola hepatica*.[24,25]

Studies for the fast-track discovery of novel paramphistomicides, antimalarial, and antibacterial compounds were also conducted with this theoretical approach.[26−29]

Later, promising results in the field of bioinformatics were found in the modeling of the interaction between drugs and the HIV Ψ-RNA packaging region using the TOMOCOMD−CANAR (***C***omputed-***A***ided ***N***ucleic ***A***cid ***R***esearch) approach.[30,31] Finally, an alternative formulation of our approach for structural characterization of proteins was carried out recently.[32,33] This extended method [TOMOCOMD−CAMPS (***C***omputed-***A***ided ***M***odelling in ***P***rotein ***S***cience)] was used to encompass protein stability studies, specifically how alanine substitution mutation on Arc repressor wild-type protein affects protein stability, by means of a combination of protein linear or quadratic indices (macromolecular fingerprints) and statistical (linear and nonlinear model) methods.[32,33]

In the present work, the TOMOCOMD−CARDD strategy is used to find quantitative models that allow the discrimination of antimalarial compounds from inactive ones in a rational way, using nonstochastic and stochastic quadratic indices. A virtual screening for the search of new lead compounds with a novel action mechanism is performed for the case of Ras FTPase inhibitors with antimalarial activity. Finally, we present the design, synthesis, and in vitro evaluation against two *Plasmodium falciparum* strains of synthetic intermediaries of quinolinic compounds, as a starting point for the development of new inexpensive antimalarials.

2. THEORETICAL FRAMEWORK

The theoretical scaffold of the TOMOCOMD−CARDD's molecular descriptors family was split in the present paper into two parts: one describing the mathematical features of nonstochastic fingerprints and the other one related to the stochastic quadratic indices.

2.1. Nonstochastic Quadratic Fingerprints. Implemented in the CARDD package of the TOMOCOMD software, the atom, atom-type, and total nonstochastic quadratic fingerprints can be calculated from both the molecular pseudograph's atom adjacency matrix and the molecular vector of small-to-medium-sized organic compounds. The general principles of these quadratic indices have been explained in some detail elsewhere.[14,18,20−23,25,26] However, an overview of this approach will be given.

For a given molecule composed of n atoms, the "molecular vector" (X) is constructed and the kth total quadratic indices, $q_k(x)$, are calculated as quadratic forms, as shown in eq 1

$$q_k(x) = \sum_{i=1}^{n}\sum_{j=1}^{n} {}^{k}a_{ij}x_i x_j \tag{1}$$

where n is the number of atoms in the molecule and x_1-x_n are the coordinates or components of the "molecular vector" (X) in a system of canonical basis vectors of $\mathscr{R}^n$. The components of the molecular vector are numerical values, which can be considered as weights (atom labels) for the vertices of the pseudograph. Certain atomic properties (electronegativity, atomic radii, etc.) can be used with this propose. In this work, the Pauling electronegativities are selected as atom weights.[34]

The coefficients ${}^{k}a_{ij}$ are the elements of the kth power of the symmetrical square matrix $\mathbf{M}$(G) of the molecular pseudograph (G) and are defined as follows:

$$\begin{aligned} a_{ij} &= P_{ij} \text{ if } i \neq j \text{ and } \exists\, e_k \in \mathrm{E(G)} \\ &= L_{ii} \text{ if } i = j \\ &= 0 \text{ otherwise} \end{aligned} \tag{2}$$

where E(G) represents the set of edges of G, P_{ij} is the number of edges (bonds) between vertices (atoms) v_i and v_j, and L_{ii} is the number of loops in v_i (see Table 1).

Equation 1 for $q_k(x)$ can be written as the single matrix equation

$$\mathbf{q}_k(x) = \mathbf{X}^{t}\mathbf{M}^{k}\mathbf{X} \tag{3}$$

where $\mathbf{X}$ is a column vector (a $n \times 1$ matrix), $\mathbf{X}^{t}$ is the transpose of $\mathbf{X}$ (a $1 \times n$ matrix), and $\mathbf{M}^k$ is the kth power of the matrix $\mathbf{M}$ of the molecular pseudograph G (mathematical quadratic form's matrix).

In addition to total quadratic indices, computed for the whole molecule, local-fragment (atom and atom-type) formalisms can be developed. These descriptors are termed local quadratic indices, $q_{kL}(x)$.[14,18,20−23,25,26] The definition of these descriptors is as follows:

$$q_{kL}(x) = \sum_{i=1}^{m}\sum_{j=1}^{m} {}^{k}a_{ijL}x_i x_j \tag{4}$$

where m is the number of atoms of the fragment of interest and ${}^{k}a_{ijL}$ is the element of row "i" and column "j" of the matrix $\mathbf{M}^{k}_{L}$. This matrix is extracted from the $\mathbf{M}^k$ matrix and contains information referred to the vertices (atoms) of the specific molecular fragments and of the molecular environment. The matrix $\mathbf{M}^{k}_{L} = [{}^{k}a_{ijL}]$ with elements ${}^{k}a_{ijL}$ is defined as follows:

$$\begin{aligned} {}^{k}a_{ijL} &= {}^{k}a_{ij} \quad \textit{if} \text{ both } v_i \text{ and } v_j \text{ are atoms} \\ &\qquad \text{contained within the molecular fragment} \\ &= {}^{1}/_{2}\,{}^{k}a_{ij} \quad \textit{if } v_i \text{ or } v_j \text{ is an atom contained within the} \\ &\qquad \text{molecular fragment but not both} \\ &= 0 \quad \text{otherwise} \end{aligned} \tag{5}$$

These local analogues can also be expressed in matrix form by the expression

$$\mathbf{q}_{kL}(x) = \mathbf{X}^{t}\mathbf{M}^{k}_{L}\mathbf{X} \tag{6}$$

Notice that the above scheme follows the spirit of a Mulliken population analysis.[35] Also note 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 matrix $\mathbf{M}^k$ can be partitioned into Z local matrices $\mathbf{M}^{k}_{L}$ ($L = 1, ..., Z$), and the kth power of matrix $\mathbf{M}$ is exactly the sum of the kth power of the local Z matrices. In this way, the total quadratic indices are the sum of the quadratic indices of the Z molecular fragments:

$$q_k(x) = \sum_{L=1}^{Z} q_{kL}(x) \tag{7}$$

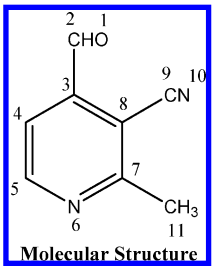
Atom and atom-type quadratic fingerprints are specific cases of local quadratic indices. In this sense, the kth atom-type quadratic indices are calculated by adding the kth atom quadratic indices for all atoms of the same type in the molecule.

In the atom-type quadratic indices formalism, each atom in the molecule is classified into an atom type (fragment), such as heteroatoms, hydrogen bonding (H bonding) to heteroatoms (O, N, and S), halogen atoms, aliphatic carbon chain, aromatic atoms (aromatic rings), and so on. For all data sets, considering those with a common molecular scaffold as well as those with a diverse structure, the kth atom-type quadratic indices provide important information.

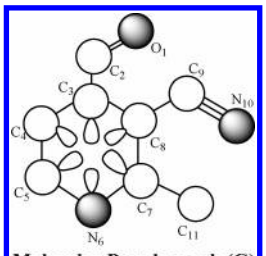
2.2. Atom, Atom-Type, and Total Stochastic Quadratic Fingerprints. Notice that the mathematical quadratic form matrices, $\mathbf{M}^k$, are graph-theoretical electronic-structure models, like the "extended Hückel" model. The $\mathbf{M}^1$ matrix considers all valence-bond electrons (σ and π networks) in one step, and their power k ($k = 0, 1, 2, 3, ...$) can be considered as an interacting-electronic chemical-network model in k steps. This model can be seen as an intermediate one between the quantitative quantum-mechanical Schrödinger equation and classical chemical bonding ideas.[38]

Recently, our research group developed a new method based on the Markov chain theory, which has been successfully used in QSPR and QSAR studies.[13,37,39] This approach also describes changes in the electron (stochastic) distribution and vibrational decay over time throughout the molecular backbone using the Markov chain formalism.

Table 1. Calculation of $\mathbf{M}^k$(G) and $\mathbf{S}^k$(G) for 2-formyl-6-methyl-benzonitrile, When k Varies from 0 to 2 and i Is a Specific Atom in the Molecule



Molecular Structure



Molecular Pseudograph (G)

a_{ij}	O_1	C_2	C_3	C_4	C_5	N_6	C_7	C_8	C_9	N_{10}	C_{11}	${}^k\delta_i$	O_1	C_2	C_3	C_4	C_5	N_6	C_7	C_8	C_9	N_{10}	C_{11}
						$\mathbf{M}^0$(G)												$\mathbf{S}^0$ (G)					
O_1	1	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0
C_2	0	1	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0
C_3	0	0	1	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0
C_4	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0
C_5	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0
N_6	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0
C_7	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0
C_8	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0
C_9	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	1	0	0
N_{10}	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	1	0
C_{11}	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	1
						$\mathbf{M}^1$(G)												$\mathbf{S}^1$(G)					
O_1	0	2	0	0	0	0	0	0	0	0	0	2	0	1	0	0	0	0	0	0	0	0	0
C_2	2	0	1	0	0	0	0	0	0	0	0	3	0.66	0	0.33	0	0	0	0	0	0	0	0
C_3	0	1	1	1	0	0	0	1	0	0	0	4	0	0.25	0.25	0.25	0	0	0	0.25	0	0	0
C_4	0	0	1	1	1	0	0	0	0	0	0	3	0	0	0.33	0.33	0.33	0	0	0	0	0	0
C_5	0	0	0	1	1	1	0	0	0	0	0	3	0	0	0	0.33	0.33	0.33	0	0	0	0	0
N_6	0	0	0	0	1	1	1	0	0	0	0	3	0	0	0	0	0.33	0.33	0.33	0	0	0	0
C_7	0	0	0	0	0	1	1	1	0	0	1	4	0	0	0	0	0	0.25	0.25	0.25	0	0	0.25
C_8	0	0	1	0	0	0	1	1	1	0	0	4	0	0	0.25	0	0	0	0.25	0.25	0.25	0	0
C_9	0	0	0	0	0	0	0	1	0	3	0	4	0	0	0	0	0	0	0	0.25	0	0.75	0
N_{10}	0	0	0	0	0	0	0	0	3	0	0	3	0	0	0	0	0	0	0	0	1	0	0
C_{11}	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0
						$\mathbf{M}^2$(G)												$\mathbf{S}^2$(G)					
O_1	4	0	2	0	0	0	0	0	0	0	0	6	0.66	0	0.33	0	0	0	0	0	0	0	0
C_2	0	5	1	1	0	0	0	1	0	0	0	8	0	0.625	0.125	0.125	0	0	0	0.125	0	0	0
C_3	2	1	4	2	1	0	1	2	1	0	0	14	0.143	0.071	0.287	0.143	0.071	0	0.071	0.143	0.071	0	0
C_4	0	1	2	3	2	1	0	1	0	0	0	10	0	0.1	0.2	0.3	0.2	0.1	0	0.1	0	0	0
C_5	0	0	1	2	3	2	1	0	0	0	0	9	0	0	0.111	0.222	0.333	0.222	0.111	0	0	0	0
N_6	0	0	0	1	2	3	2	1	0	0	1	10	0	0	0	0.1	0.2	0.3	0.2	0.1	0	0	0.1
C_7	0	0	1	0	1	2	4	2	1	0	1	12	0	0	0.083	0	0.083	0.166	0.333	0.166	0.083	0	0.083
C_8	0	1	2	1	0	1	2	4	1	3	1	16	0	0.063	0.125	0.063	0	0.063	0.125	0.25	0.063	0.188	0.063
C_9	0	0	1	0	0	0	1	1	10	0	0	13	0	0	0.077	0	0	0	0.077	0.077	0.769	0	0
N_{10}	0	0	0	0	0	0	0	3	0	9	0	12	0	0	0	0	0	0	0	0.25	0	0.75	0
C_{11}	0	0	0	0	0	1	1	1	0	0	1	4	0	0	0	0	0	0.25	0.25	0.25	0	0	0.25

The present approach is based on a simple model for the intramolecular (stochastic) movement of all valence-bond electrons. Let us consider a hypothetical situation in which a set of atoms is free in space at an arbitrary initial time (t_0). At this time, the electrons are distributed around atomic nuclei. Alternatively, these electrons can be distributed around cores in discrete intervals of time t_k. In this sense, the electron at an arbitrary atom i can move to other atoms at different discrete time periods t_k (k = 0, 1, 2, 3...) throughout the chemical-bonding network.

The kth stochastic molecular pseudograph's atom adjacency matrix [$\mathbf{S}^k$(G)] can be obtained from $\mathbf{M}^k$. Here, $\mathbf{S}^k$(G) = $\mathbf{S}^k$ = [${}^k s_{ij}$] is a squared table of order n (n = number of atoms), and the elements ${}^k s_{ij}$ are defined as follows:

$$ {}^k s_{ij} = \frac{{}^k a_{ij}}{{}^k \sum_i} = \frac{{}^k a_{ij}}{{}^k \delta_i} \qquad (9) $$

where ${}^k a_{ij}$ are the elements of the kth power of $\mathbf{M}$ and the sum ($\sum$) of the ith row of $\mathbf{M}^k$ is named the k-order vertex degree of atom i, ${}^k\delta_i$. The kth s_{ij} elements are the transition probabilities with which the electrons move from atom i to atom j in the discrete time period t_k (step by step). Notice that the kth elements s_{ij} take into consideration the information of the molecular topology in step k throughout the chemical-bonding (σ and π) network. For instance, the ${}^2 s_{ij}$ values can distinguish between the hybrid states of atoms in bonds. In this sense, it can clearly be seen from Table 1 that electrons will have a higher probability of returning to the sp N atom [$p(N_{10}) = 0.75$] than to the sp^2 N atom [$p(N_6) = 0.33$] in t_2. A similar behavior can be observed among the different hybrid states of the C atoms in the molecule of 2-formyl-6-methyl-benzonitrile (see Table 1): C sp^3 [$p(C_{11}) = 0.25$]; C sp^2 [$p(C_2) = 0.625$]; C $sp^2{}_{arom}$ [$p(C_3) = 0.285$, $p(C_4) = 0.3$, $p(C_5) = 0.33$, $p(C_7) = 0.33$, $p(C_8) = 0.25$]; and C sp [$p(C_9) = 0.769$]. This is a logical result if the electronegativity scale of these hybrid states is taken into account. The kth total and local stochastic quadratic indices,

${}^{s}q_k(x)$, are calculated in the same way that the nonstochastic quadratic indices are, but using the *k*th stochastic molecular pseudograph's atom adjacency matrix, $\mathbf{S}^k$(G), as the mathematical quadratic forms' matrix.

3. MATERIALS AND METHODS

3.1. Computational Methods: TOMOCOMD−CARDD Approach. Molecular fingerprints were generated by means of the interactive program for molecular design and bioinformatic research TOMOCOMD.[15] The CARDD module′ was selected for drawing all structures and for the computation of nonstochastic and stochastic indices.

The main steps for the application of this method in QSAR/QSPR and for drug design can be briefly summarized as follows:

1. Drawing of the molecular pseudographs for each molecule of the data set, using the drawing mode.

2. Selection of appropriate weights in order to differentiate molecular atoms.

3. Computation of the total and local (atom and atom-type) quadratic indices of the molecular pseudograph's atom adjacency matrix using the calculation mode. Atomic properties and family descriptors are also selected.

4. Development of a QSPR/QSAR equation by using several multivariate analytical techniques, such as multilinear regression analysis, neural networks, linear discriminant analysis (LDA), and so on. A quantitative relationship between an activity *A* and the quadratic fingerprints is generated, having for instance, the following appearance:

$$A = a_0q_0(x) + a_1q_1(x) + a_2q_2(x) + ... + a_kq_k(x) + c \quad (10)$$

where *A* is the measured activity, $q_k(x)$ are the *k*th total quadratic indices, and the a_k's are the coefficients obtained by the linear regression analysis.

5. Test of the robustness and predictive power of the QSPR/QSAR equation by using internal [leave-*one*-out (LOO) and LGO cross validation] and external (using a test set and an external predicting set) validation techniques.

The following descriptors were calculated in this work:

(i) *k*th total quadratic indices not considering and considering H atoms in the molecular pseudograph (G) [$q_k(x)$ and $q_k^{H}(x)$, respectively].

(ii) *k*th local (atom type = heteroatoms: S, N, O) quadratic indices not considering and considering H atoms in the molecular pseudograph (G) [$q_{kL}(x_E)$ and $q_{kL}{}^{H}(x_E)$, respectively]. These local descriptors are putative H-bonding acceptors.

(iii) *k*th local (atom type = H atoms bonding to heteroatoms: S, N, O) quadratic indices considering H atoms in the molecular pseudograph (G) [$q_{kL}{}^{H}(x_{E-H})$]. These local descriptors are putative H-bonding donors.

The *k*th stochastic total [${}^{s}q_k(x)$ and ${}^{s}q_k{}^{H}(x)$] and local [${}^{s}q_k(x_E)$, ${}^{s}q_k{}^{H}(x_E)$, and ${}^{s}q_k{}^{H}(x_{E-H})$] quadratic indices were also computed.

3.2. Data Set. A large data set of 1562 organic chemicals having a great structural variability—597 of them antimalarial agents[2,7−9,40−82] and the rest inactive ones[41,82] (965 compounds having other clinical uses, such as antivirals, sedative/hypnotics, diuretics, anticonvulsivants, hemostatics, oral hypoglycemics, antihypertensives, antihelminthics, anticancer compounds, and so on)—was selected.

The data set of active compounds was chosen considering a representation of most of the different structural patterns and action modes for the case of compounds with antimalarial activity. It includes (1) alkaloidal and synthetic quinoline-based antimalarial drugs that involve the blockage of the function of the food vacuole (4- and 8-aminoquinolines,[9,70] peptide derivatives,[52] dimeric quinolines,[47,49] and other compounds such as indolo[3,2-*c*]quinolines[7] and methylene blue derivatives), (2) peptide (fluoromethyl ketone peptide derivatives) and nonpeptide (phenothiazines and chalcones) falcipain-cysteine protease inhibitors,[42,45] (3) peptide and nonpeptide inhibitors of malarials aspartyl protease plasmepsin II,[48] (4) agents interfering with *P. falciparum* phospholipid metabolism (primary, secondary, tertiary amines and quaternary ammonium and bisammonium salts),[69] (5) antimalarials that have the ability to inhibit electron transport processes and respiratory systems by acting as ubiquinone antagonists (hydroxynaphthoquinones such as atovaquone),[40] (6) selective inhibitors of lactate dehydrogenase from the malaria parasite (derivatives of the sesquiterpene 8-deoyhemigossylic acid),[40] (7) antimalarial chemicals that act by selectively inhibiting malarial dihydrofolate reductase−thymidylate synthase (pyrimethamine and it is analogues),[8] (8) antiparasitic agents affecting DNA topoisomerases (e.g., anticancer acridines),[53] and (9) artemisinin-type antimalarials and other simple bicyclic and tetracyclic endoperoxides (including lactone ring-open analogues the trioxanes).[2−5,44,51,54−65,66−68,72] Other compounds for which there have not been found or defined a specific mode of action but which have been reported as antimalarial agents were also included.[41,50,82]

Later, two k-means cluster analyses (k-MCA) were performed for active and inactive series of compounds, to split the data set (1562 organic chemicals) into training and predicting series.[83,84] All cases were processed using k-MCA in order to design training and predicting data series in a "rational" way.

3.3. Chemometric Methods. k-Means Cluster Analysis (k-MCA). The statistical software package STATISTICA[85] was used to develop the k-MCA.[85] The number of members in each cluster and the standard deviation of the variables in the cluster (kept as low as possible) were taken into account, to have an acceptable statistical quality of data partitions in the clusters. The values of the standard deviation (SS) between and within clusters, of the respective Fisher ratio and their *p* level of significance, were also examined.[83,84]

Linear Discriminant Analysis (LDA). LDA was carried out with the STATISTICA software.[85] The considered tolerance parameter (proportion of variance that is unique to the respective variable) was the default value for minimum acceptable tolerance, which is 0.01. A forward stepwise search procedure was fixed as the strategy for variable selection. The principle of parsimony (Occam's razor) was taken into account as a strategy for model selection. In connection, we selected the model with a high statistical significance but having as few parameters (a_k) as possible.

The quality of the models was determined by examining Wilks' λ parameter (*U* statistic), the square Mahalanobis distance (D^2), the Fisher ratio (*F*), and the corresponding *p*

level [$p(F)$] as well as the percentage of good classification in the training and test sets. Models with a proportion between the number of cases and variables in the equation lower than 5 were rejected.

The biological activity was codified by a dummy variable "***Class***". This variable indicates the presence of either an active compound (***Class*** = 1) or an inactive compound (***Class*** = −1). The classification of cases was performed by means of the posterior classification probabilities. By using the models, one compound can then be classified as active, if $\Delta P\% > 0$, being $\Delta P\% = [P(\text{Active}) - P(\text{Inactive})]100$, or as inactive otherwise. P(Active) and P(Inactive) are the probabilities with which the equations classify a compound as active or inactive, respectively.

The statistical robustness and predictive power of the obtained model was assessed using a prediction (test) set. Also, a leave-group-out (LGO) cross-validation strategy was carried out.[86] In this case, 10% of the data set was used as group size; that is, groups including 10% of the training data set are left out and predicted for 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 the robustness and stability of the obtained models.

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 (MCC) in the training and test sets permitted the assessment of the model.[88]

Orthogonalization of Descriptors. The Randić method of orthogonalization was used.[89−95] As a first step, an appropriate order of orthogonalization was considered following the order with which the variables were selected in the forward stepwise search procedure of the statistical analysis.[95] The first variable (V_1) is taken as the first orthogonal descriptor ${}^1O(V_1)$, and the second one (V_2) is orthogonalized with respect to it [${}^2O(V_2)$]. The residual of its correlation with ${}^1O(V_1)$ is that part of the descriptors V_2 not reproduced by ${}^1O(V_1)$. Similarly, from the regression of V_3 versus ${}^1O(V_1)$, the residual is the part of V_3 that is not reproduced by ${}^1O(V_1)$, and it is labeled ${}^1O(V_3)$. The orthogonal descriptor ${}^3O(V_3)$ is obtained by repeating this process in order to also make it orthogonal to ${}^2O(V_2)$. The process is repeated until all variables are completely orthogonalized, and the orthogonal variables are then used to obtain the new model.

3.4. Chemistry. IR spectra were recorded with a FTIR−BOMEM spectrometer using KBr disks for solids or NaCl cells for liquids (υ in cm^{-1}). ^1H NMR and ^{13}C NMR spectra were recorded on a Bruker ADPX-300 (300 mHz) using $CDCl_3$ as the solvent. The calibration of the spectra was carried out on tetramethylsilane (internal ^1H) and $CDCl_3$ (^{13}C), signals δ ^1H (TMS) = 0 and δ ^{13}C ($CDCl_3$) = 77.0. Chloroquine diphosphate was supplied by "Fundação para o Remédio Popular" (Brazil). All solvents were previously dried and purified before use, according to standards established in the literature.[96,97]

3.5. Determination of in Vitro Antiplasmodial Activity. An in vitro antiplasmodial evaluation was performed by using the susceptibility microtechnique.[98] Two strains of *P. falciparum*, K1-chloroquine resistant and Palo Alto-chloroquine sensitive, kindly provided by the WHO Registry of Standard Strains of Malaria Parasites at the University of Edinburgh, were continuously maintained in culture and used in these assays.[99] The parasites' freezing and thawing procedures were carried out as already reported.[100] The parasites were cultivated to 5% hematocrit in RPMI 1640 medium with 25 mM HEPES, 21 mM sodium bicarbonate, 370 μM hypoxantine, 40 μg/mL gentamycin, and 10% human A^+ or O^+ serum provided by Fundação Pró-Sangue/Hemocentro de São Paulo. Washed human O^+ erythrocytes were added to the culture. Synchronization was obtained by treatment with D-sorbitol when the parasites were predominantly in the young trophozoite stage.[101] Stock solutions of the compounds (1 000 pmol/100 μL of ethanol) were used to prepare different concentrations (1, 2, 4, 6, 8, 16, 32, and 100 pmol/well) in aqueous solution. A stock solution of chloroquine diphosphate (1 000 pmol/100 μL in water) was used to prepare a series of concentrations (1, 2, 4, 6, 8, 16, and 32 pmol/well) to check the sensitivity of the samples. Flat-bottomed microtiter plates were dosed adding 100 μL of each concentration per well. The plates were dried at 37 °C and stored at 4 °C. An aliquot of 100 μL of the culture with a parasitemia between 0.5 and 1.0% and the parasites in the young trophozoite stage was added to each well of the microtiter plates. A control without the compound and a sensitivity test to chloroquine were performed in parallel. Microplates were incubated in a candle jar with a gas mixture of 3% CO_2, 5% O_2, and 92% N_2 and maintained at 37 °C for 24−36 h. Giemsa-stained thick blood smears were prepared from each well when controls showed the presence of schizonts by optical microscopy. The number of schizonts was counted per 200 asexual parasites, and the tests were considered valid when this number was equal or superior to 10%. The minimum inhibitory concentration (MIC) of each compound was defined by the lowest concentration that completely inhibited the schizont maturation.

4. RESULTS AND DISCUSSION

4.1. Training and Test Set Design through k-MCA. It is well-known that the quality of a classification model is highly dependent on the quality of the selected data set. The most critical aspect for constructing the training set is to warrant enough molecular diversity for it. Taking this into account, we selected a large data set of 1562 organic chemicals having a great degree of structural variability. The antimalarials considered on this study are representative ones of families with different action modes and diverse structural patterns. Figure 1 shows a representative sample of such active compounds.

The inactive group included antivirals, sedative/hypnotics, diuretics, anticonvulsivants, hemostatics, oral hypoglycemics, antihypertensives, antihelminthics, anticancer compounds, and so on, guaranteeing great structural variability. However, the declaration of these compounds as "inactive" antimalarials per se does not guarantee that antimalarial side effects do not exist for some of those organic-chemical drugs that have been left undetected so far. This problem can be reflected in the results of classification for the series of inactive chemicals.

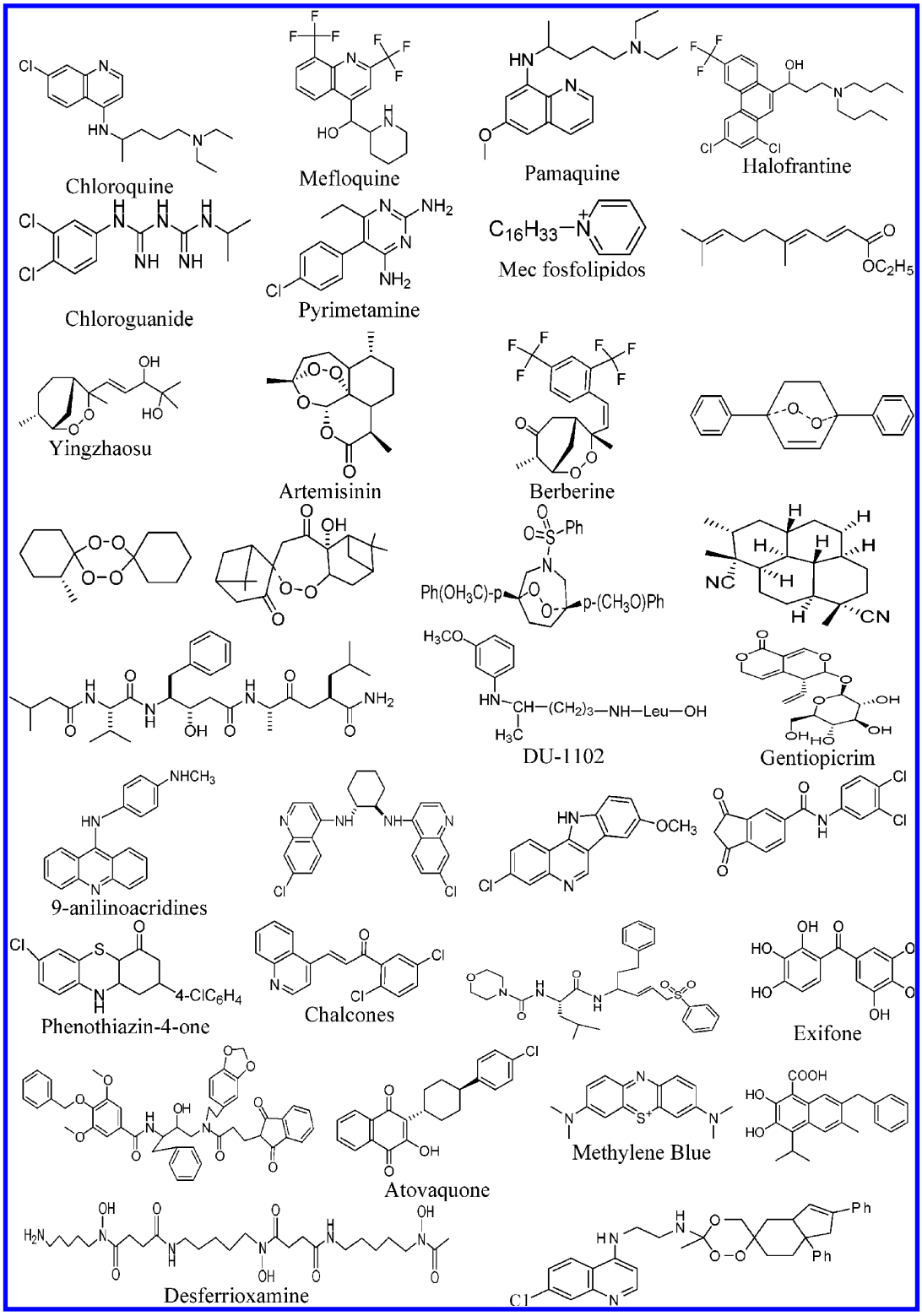


Figure 1. Random, but not exhaustive, sample of the molecular families of antimalarial agents studied here.

To split the whole group into two data sets (training and predicting ones), two *k*-MCAs[83,84] were performed for active and inactive compounds, respectively. The main idea of this procedure consists of making a partition of either active or inactive series of chemicals in several statistically representative classes of compounds. This procedure ensures that any chemical class (as determined by the clusters derived from k-MCA) will be represented in both compounds' series. This "rational" design of training and predicting series allowed us to design both sets that are representative of the whole "experimental universe".

We first carried out a k-MCA with active compounds, and afterward, we did so with inactive ones. A first k-MCA (k-MCA I) split the antimalarials into 20 clusters with 33, 18, 29, 29, 21, 59, 46, 57, 37, 16, 9, 35, 24, 55, 17, 22, 25, 34, 13, and 18 members. Additionally, the inactive compound

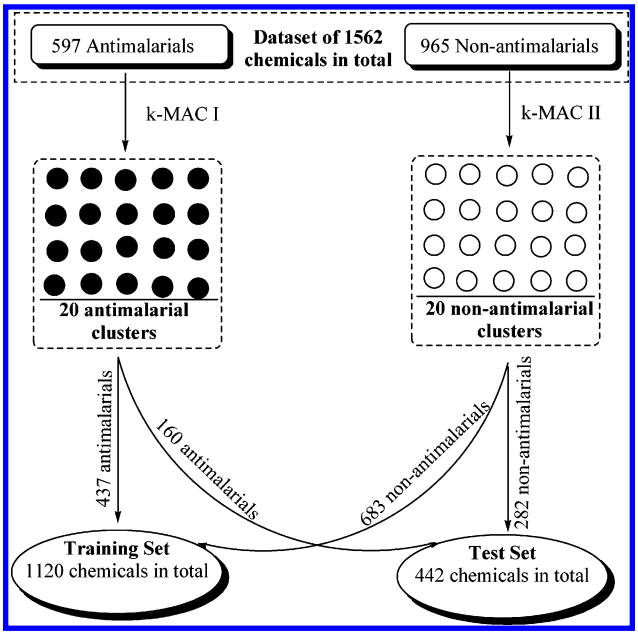


Figure 2. General algorithm used to design training and test sets throughout *k*-MCA.

series was also partitioned into 20 clusters (k-MCA II) with 58, 26, 78, 26, 48, 64, 60, 53, 80, 72, 46, 64, 41, 68, 58, 25, 4, 22, 23, and 49 members.

Afterward, the selection of the training and prediction sets was performed by taking, in a random way, compounds belonging to each cluster. From these 1562 compounds, 1120 were chosen at random to form the training set, 437 of them being actives and 683 being 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 antimalarial activity but also with novel modes of action. This will be well-illustrated in this paper in a virtual experiment for lead compound generation.

The remaining subseries composed of 160 antimalarials and 282 compounds with different biological properties were prepared as test sets for the external cross validation of the models. These compounds were never used in the development of the classification models. Figure 2 graphically illustrates the above-described procedure where two independent cluster analyses (one for active and the other for inactive chemicals) were performed to select a representative sample for the training and test sets.

The *k*th total and atom-type nonstochastic quadratic indices were used, with all variables showing p levels < 0.05 for the Fisher test. The results are depicted in Table 2.

From the k-MCA, it can be concluded that the structural diversity of several up-to-date known antimalarials (as codified by TOMOCOMD−CARDD descriptors) may be described at least by 20 statistically homogeneous clusters of chemicals.

4.2. Developing Classification Functions. Despite the number of existing chemometric techniques to find discriminant functions, such as SIMCA or neural networks, we selected LDA, given the simplicity of the method. The use of LDA in rational drug design has been extensively reported by different authors.[12−14,18−20,22−30] It being the key of the present study, we developed two classification functions using topological descriptors computed with the TOMO-

Table 2. Main Results of the *k*-MCA for Antimalarial and Nonantimalarial Chemicals

	analysis of variance			
total and atom-type quadratic indices	between SS[a]	within SS[b]	Fisher ratio (F)	p level[c]
Antimalarial Agent Clusters (*k*-MCA I)				
$q_0(x)$	380.25	38.27	301.77	0.00
$q_1(x)$	409.21	49.23	252.41	0.00
$q_{2L}(x_E)$	330.71	49.54	202.71	0.00
$q_0^H(x)$	303.94	59.04	156.35	0.00
$q_2^H(x)$	347.62	48.79	216.35	0.00
$q_{1L}^H(x_E)$	325.50	44.26	223.32	0.00
$q_{2L}^H(x_E)$	350.39	41.50	256.43	0.00
$q_{1L}(x_{E-H})$	332.62	118.54	85.21	0.00
$q_{15L}(x_{E-H})$	445.03	97.93	138.01	0.00
Nonantimalarial Agents Clusters (*k*-MCA II)				
$q_0(x)$	273.57	41.32	329.26	0.00
$q_1(x)$	367.53	54.29	336.73	0.00
$q_{2L}(x_E)$	474.27	71.02	332.13	0.00
$q_0^H(x)$	269.53	63.90	209.78	0.00
$q_2^H(x)$	378.68	52.31	360.02	0.00
$q_{1L}^H(x_E)$	398.26	64.82	305.59	0.00
$q_{2L}^H(x_E)$	443.78	65.90	334.91	0.00
$q_{1L}(x_{E-H})$	430.15	125.61	170.32	0.00
$q_{15L}(x_{E-H})$	271.09	122.98	109.64	0.00

[a] Variability between groups. [b] Variability within groups. [c] Level of significance.

COMD−CARDD software.[15] These linear models are given below together with their statistical parameters:

$$\textit{Class} = -10.059 - 0.088\,44q_0(x) + 0.070\,85q_1(x) + 0.189\,07q_0^H(x) - 0.0256q_2^H(x) + 0.0528q_{2L}(x_E) + 0.198\,49q_{1L}^H(x_E) - 0.099\,13q_{2L}^H(x_E) - 0.198\,16\,q_{1L}(x_{E-H}) + 2.658 \times 10^{-8}q_{15L}(x_{E-H})$$

$$N = 1120;\ \lambda = 0.32;\ D^2 = 8.8;\ F(9, 1110) = 258.32;\ p < 0.0001 \quad (11)$$

$$\textit{Class} = -8.7734 + 0.7734^s q_0^H(x) + 0.840\,22^s q_1^H(x) - 1.205\,67^s q_2^H(x) + 0.296\,27^s q_{1L}^H(x_E) - 0.3805^s q_3^H(x) - 0.1833^s q_{1L}(x_E) + 2.3858^s q_0(x_{E-H}) - 1.0558^s q_{1L}(x_{E-H}) - 1.1887^s q_{2L}(x_{E-H}) - 0.7662^s q_{3L}(x_{E-H})$$

$$N = 1120;\ \lambda = 0.35;\ D^2 = 7.7;\ F(10,1109) = 203.11;\ p < 0.0001 \quad (12)$$

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

The classification of the cases was performed by means of the posterior classification probabilities. This is the probability of the respective case belonging to a particular group (active or inactive), and it is proportional to the Mahalanobis distance. The posterior probability is the probability, based on our knowledge of the values of others variables, of the respective case belonging to a particular group.

Model 11, which includes nonstochastic indices, correctly classified 94.73% of the compounds in the training dataset, misclassifying only 59 compounds out of a total of 1120. The percentage of false actives in this data set was only

Table 3. Global Results of the Classification of Compounds in the Training and Test Sets

	Matthews' correlation coefficient	accuracy "Q_{Total}" (%)	sensitivity "hit rate" (%)	specificity (%)	false positive rate "false alarm rate" (%)
		Nonstochastic Descriptors (eq 11)			
training set	0.889	94.73	92.2	94.2	3.6
test set	0.866	93.89	88.1	94.6	2.8
		Stochastic Descriptors (eq 12)			
training set	0.855	93.13	90.2	92.1	4.9
test set	0.857	93.44	89.4	92.3	4.2

3.66%; that is, 25 inactive compounds were classified as actives out of 683 cases. Conversely, 34 compounds from the group of 437 actives were misclassified as inactive ones (7.78% misclassification).

The statistical analysis of model 12 showed similar results. In this case, the overall accuracy of the model was 93.13%. Only a 4.98% misclassification for the inactive group was observed (34 inactive compounds were classified as active ones out of a total of 683). In this case, 43 compounds out of 437 (9.84%) were false inactives.

4.2.1. Model Validation. The classification of all compounds in the complete training dataset provides some assessment of the goodness of fit of the models, but it does not provide a thorough criterion of how the models can predict the biological properties of new compounds. To assess such predictive power, the use of an external test set is essential.[86,87] In this sense, the activity of the compounds in such a set was predicted with the two obtained discrimination functions. The overall accuracy for this group was 93.89% (27/442) and 93.44% (29/442) using models 11 and 12, respectively. Taking into account the number of compounds used in the external test set, we can see that model 11 correctly classifies 97.16% (274/282) of the inactives and 88.13% (141/160) of the actives, whereas model 12 correctly classifies 95.74% (270/282) of the inactives and 89.38% (143/160) of the considered antimalarials. It can be seen that the number of misclassified inactive compounds is relatively low for both models. This is a desirable condition to consider that a model is adequate, taking into account that this number represents inactive compounds that will be sent to biological assays and, in this way, represents a loss of time and resources.[12]

The results of a global classification of compounds, in both training and external prediction sets, are shown in Table 3. This table also lists most of the parameters commonly used in medical statistics (accuracy, sensitivity, specificity, and false positive rate) and the MCC for both obtained models.[88] These models, eqs 11 and 12, showed high MCCs of 0.89 (0.87) and 0.86 (0.86) in the training (test) sets, respectively. Whereas 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, MCC 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.[88]

A second experiment, considering a LGO strategy, was carried out for both models as an internal validation procedure.[86] This method systematically removes a group of data points at a time from the data set. A QSAR/QSPR model is then constructed on the basis of this reduced data set and subsequently used to predict the removed group. In the present study, 10% of the data set was removed at each time. This procedure is repeated until a complete predictions set is obtained. Good results in this experiment are often considered as a proof of the high predictive ability of the models. However, this assumption is not always correct, and it can be that there is a poor correlation between the good LOO results and the high predictive ability of QSAR/QSPR models.[86,87] Thus, the good behavior of models in an LOO procedure appears to be the necessary but not the sufficient condition for the models to have a high predictive power.

Table 4. Predictivity Based on the Use of 10 Randomly Selected Subsets (LGO cross validation) of LDA Models

	% global good classification	
group	eq 11	eq 12
1	96.43	91.97
2	95.54	94.64
3	83.93	83.93
4	91.07	92.86
5	96.43	97.32
6	93.75	92.86
7	97.32	95.54
8	98.21	99.11
9	96.43	92.86
10	90.18	86.60
overall mean	93.93	92.77
standard deviation	4.39	4.58

The overall mean of the correct classification for this process for eqs 11 and 12 were 93.93% and 92.77%, respectively. For a 10% full leave-out cross-validation procedure, this level of cross-validated classification is a good indication of the robustness and stability of the obtained models. The results of the LGO procedure are shown in Table 4.

In summary, the calculation of percentages of good classification in the training and external data sets and an internal cross-validation procedure permitted us to carry out the assessment of the models.

In Tables 5 and 6, the results of classification using models 11 and 12 for some active and inactive compounds in the training and external prediction data sets are shown. The complete set of compounds in these series, as well as their classification, using both models is given as Supporting Information.

4.2.2. Orthogonalization of Descriptors. A close inspection of the molecular descriptors included in both LDA-based QSAR models showed that several of these fingerprints are strongly interrelated to each other. In Table 7, we give the correlation coefficients of the molecular descriptors in models 11 and 12.

The orthogonalization process of molecular descriptors was introduced by Randić several years ago as a way to improve

Table 5. Classification of Active Compounds in the Training and Test Sets by the Use of LDA-QSAR Models Developed Using Nonstochastic (eq 11) and Stochastic (eq 12) Quadratic Indices

name	$\Delta P\%^a$	$\Delta P\%^b$	name	$\Delta P\%^a$	$\Delta P\%^b$
Training Active Group					
cinchonine	−78.57	−82.98	chalcone	−94.93	−37.68
hydroxychloroquine	59.02	62.34	buquinolate	98.79	99.67
CDRI 87209	−24.92	−49.49	decoquinate	99.89	99.97
WR 33,063	99.99	99.98	methyl benzoquate	66.51	85.73
WR 122,455	23.30	31.60	exifone	77.75	62.00
pamaquine	80.67	69.16	methylene blue	−99.46	−97.79
primaquine	−60.69	−11.63	WR 197236	32.16	94.63
WR 225498	93.61	98.23	piperaquine	59.26	86.15
WR 242511	95.01	98.80	Yingzhaosu A	98.50	94.47
M8506, trifluoracetylprimaquine	80.20	−45.52	Yingzhaosu C	89.38	74.08
CDRI 8053	91.76	92.46	12278 R	99.04	99.29
chloroguanide	−69.10	−17.19	desferrioxamine	100.00	99.88
chloroproguanil	−43.29	44.10	secoartemisinim	98.73	96.47
pyrimetamine	74.18	9.71	(+)-4,5-soartemisinim	98.72	96.07
trimethoprim	61.92	−14.10	licochalcone A	0.51	47.43
sulfalene	42.84	1.08	9-desmethylartemisinin	86.42	71.90
dapsone	−95.59	−49.40	axisonitrile-3	−43.14	−26.20
sulfisoxazole	89.26	76.08	berbamine	96.19	98.80
WR 99210	91.41	96.91	malagashanine	−68.27	−59.53
PS-15	80.75	97.82	berberine	−98.70	−76.10
artemisinin	96.84	94.52	10,12 peroxycalamenene	85.32	71.56
bispyroquine	98.93	92.18	simalikalactone D	97.75	99.40
fluornemethanol	91.18	99.06	gutolactone	94.41	98.02
artemisitene	92.24	90.38	lissoclinotoxin A	96.94	30.43
aminopterin	99.91	94.73	6,9-didesmethylartemisinin	86.42	71.90
B-arteether	99.46	98.20	lapinone	99.97	99.99
benzonaphthyridine 7351	99.96	99.59	gossypol	100.00	99.99
tripiperakin	90.09	92.22	CN 10443	99.97	99.99
sodium artesunate	99.69	97.85	cycloleucine	−99.68	−98.48
amquinate	49.39	80.70	nitroguanil	14.31	0.64
fluoroquine	82.84	94.42	cloguanamil	−99.30	−90.26
pentaquine	53.60	62.85	metachloridine	42.22	14.50
dabekhin	−36.27	−45.54	supazine	29.66	−88.95
methylchloroquine	90.61	90.15	cilional	6.30	−60.17
RC− 12	73.00	80.52	WR 10 488	−38.75	82.47
dimeplasmin	52.84	50.33	acedapsone	−81.07	−16.01
azamepacrine	98.96	98.75	gentiopicrim	−96.18	−23.70
mepacrine	98.00	96.87	oxychlorochin	−2.11	20.35
6-demethyl-6-difluoromethyl B-arteether	99.87	99.86	antimalarine	17.84	6.79
tetracycline	95.35	93.96	brindoxime	62.86	91.65
doxycycline	89.78	93.53	menctone	96.42	98.73
clindamycin	87.13	99.86	pyronaridine	99.89	98.90
ciprofloxacin	−97.92	21.69	quinacrine	97.88	96.80
fenozan − 50 F	98.23	99.86	amodiaquine	93.93	80.55
arteflene	97.45	90.05	aristochin	99.44	99.33
12-(3′-hydroxy-*n*-propyl)-deoxoartemisinin	99.59	98.65	2-(4-metoxybenzoyl)−1-naphthoil acid	−97.58	−93.39
oxalic bis(2-hidroxy-1-naphthylmethylene)hidrazide	62.88	−29.90			
Test Active Group					
quinine	−56.15	−61.41	iso-artemisitene	95.21	88.86
chloroquine	77.93	77.96	B-artemether	98.41	93.25
mefloquine	49.30	42.50	artelinic acid	99.76	98.31
halofrantine	93.22	98.16	methotrexate	99.54	87.52
quinocide	−61.12	−4.28	apicidin	100.00	100.00
WR 238,605	97.48	99.34	naphthol blue-black	100.00	100.00
WR182393	31.77	80.98	7,7 difluoro-B-arteether	99.87	99.84
cycloguanil	−87.33	−85.04	9-epiartemisinin	93.39	90.01
sulfadoxine	69.93	41.35	hydrolapachol	−75.01	6.04
sulfamethoxazole	71.96	42.99	atovaquone	−65.89	−24.04
clociguanil	24.93	41.50	cycloquin	100.00	99.99
nitroquine	99.76	92.53	amopyroquine	85.72	52.21
dihydroartemisinin	96.60	87.33	azithromycin	100.00	100.00
lapacol	−88.44	−45.84	brusatol	97.48	99.47
dioncophyline B	95.41	74.06	aecachinium	−11.21	15.84
tebuquine	99.09	95.08	hexalorxylol	−93.86	−95.27
strychnobrasiline	−86.33	−90.25	WR 135 403	73.05	−5.14
octanoylprimaquine	99.54	98.80	WR 226 253	30.63	77.24
norfloxacin	−98.10	4.63	CI −608	−42.29	99.67
enpiroline	60.93	48.61	endochin	79.02	76.76
refigallol	97.51	92.07	isopentachin	65.34	69.59
WR 194905	99.66	99.92	floxacrine	24.78	48.64

[a] Results of the classification of compounds obtained from eq 11 (using nonstochastic quadratic indices). [b] Results of the classification of compounds obtained from eq 12 (using stochastic quadratic indices): $\Delta P\% = [P(\text{Active}) - P(\text{Inactive})]100$ (see developing classification functions).

Table 6. Classification of Inactive Compounds in the Training and Test Sets by the Use of LDA-QSAR Models Developed Using Nonstochastic (eq 11) and Stochastic (eq 12) Quadratic Indices

name	$\Delta P\%^a$	$\Delta P\%^b$	name	$\Delta P\%^a$	$\Delta P\%^b$
		Training Inactive Group			
amantadine	−97.51	−95.25	tiazesin hidrochloride	−92.93	−95.47
dimepranol	−99.82	−99.75	imipraminoxide	−14.81	−42.24
radicinin	−95.51	−29.28	prooksen	−98.45	−95.23
halopropane	−98.18	−99.09	domazoline fumarate	−96.57	−90.78
roflurane	−99.23	−99.86	xylometazoline	−72.47	−65.27
methioflurane	−98.40	−98.21	mtrafazoline	−89.46	−82.61
cinromide	−98.55	−94.09	nigrifactin	−97.14	−96.67
zebromal	−96.66	−92.10	perastine hydrocloride	−88.19	−90.74
riodoxol	−97.44	−93.31	oxaprotine hidrochloride	−97.39	−95.37
ethoxene	−99.87	−99.72	carnidazole	−33.60	−14.30
ketoxal	−99.31	−99.55	dimetridazole	−52.78	−76.98
norflurane	−99.59	−99.94	tramazoline hydrochloride	−89.85	−85.34
tribromethanol	−99.75	−99.79	phedrazine	−96.04	−88.52
ethyl bromide	−99.89	−99.34	naphazoline	−99.53	−98.79
ethyl chloride	−99.89	−99.29	tefazoline nitrate	−93.16	11.59
moroxidine	−99.86	−99.10	N methylglucamine	−99.85	−98.33
olmidine	−99.38	−96.07	leucenol	−99.76	−97.71
calcii bromoaminoacetas	−99.98	−99.96	parnaquone	−76.44	0.04
trichlorourethan	−99.79	−99.95	antimoniato de meglumina	−99.85	−98.33
thiourethane	−99.89	−99.51	nitrofural	1.04	−65.84
noxiptiline hidrochloride	−62.75	−72.85	nafiverine	36.12	51.58
fructosum ferricum	−98.99	−99.53	acidum sultosilicum	−99.15	−29.99
ferrotrenine	−99.67	−99.57	ciprofibrate	−79.74	−87.76
chlorobutanol	−97.87	−99.04	724 oxicodona	−76.67	−40.42
vernelan	−99.95	−99.89	morfina	−77.55	−60.77
sedanfactor solution	−99.94	−99.76	petidina	−95.19	−91.54
cloretate	−90.67	−99.04	tilidina	−91.44	−88.16
benzylicum	−99.91	−99.84	chlorphenesin carbamate	−99.51	−99.05
metacaine	−96.77	−92.08	phenprobamate	−99.58	−89.15
subcutin	−97.03	−92.16	mephenesin	−99.43	−98.25
fluorotyrosinum	−98.32	−82.34	pilokarpin	−98.85	−94.36
thibenzazoline	−99.87	−99.64	eseridine	−65.82	−78.23
dimecamine	−85.18	−94.61	tyrosam	−99.68	−98.67
carbromide	−98.44	−94.47	diampromide	−56.89	−68.58
anticoman	−94.61	−29.63	meglitinide	−83.13	−70.16
furfurylurea	−90.97	−77.67	calcii mesoxalas	−99.95	−99.97
paraldehyde	−98.94	−99.33	dipipanone	14.33	10.41
hedonal	−99.43	−99.42	hydroxypethidine	−83.79	−76.79
aponal	−99.21	−99.57	ketobemidone	−79.80	−77.41
butamben	−87.02	−67.76	strychnocarpine	−88.86	−93.09
chlorphenacemide	−99.77	−99.59	medrylamine	−96.04	−97.25
phenacemide	−99.85	−99.84	metildipenhidramine	−96.35	−97.75
mercaptothiazoline	−99.95	−99.86	metaxalone	−98.75	−96.36
tiouracilo	−98.75	−99.12	carisoprodol	−85.50	−84.44
almoxatone mesilate	−96.24	−75.18	phenamacide hydrochloride	−79.81	−70.96
diazoxide	−94.38	−84.17	BA 1	−99.98	−99.90
guanabenz acetate	−98.09	−90.55	norgamem	−99.95	−99.84
pyrantel tartrate	−99.37	−98.43	EN-3638	−25.53	−49.23
aminoglutethimide	−61.43	−51.66	difenilhidramine hidrocloride	−98.26	−98.92
carbavin	−99.84	−99.89	bromazine	−56.50	−71.39
moxonidine	−67.62	−30.53	basidalin	−99.66	−96.23
mephenytoin	−97.78	−88.12	oxiniacic acid	−99.78	−98.90
primidone	−98.27	−91.89	acipimox	−98.72	−94.26
mezepine	−98.27	−97.60	ethyl glutamate	−99.38	−96.94
bifemelane hidrochloride	−97.98	−91.30	succinyl-methionine	−97.41	−97.57
glycyclamide	−52.72	−18.37	metformin	−99.96	−99.68
albutoin	−91.70	−67.98	borimamide	−99.28	−92.71
styptol	−99.14	−95.59	cianothepin	−97.91	−93.36
lodal	−99.77	−99.06	metoxepin	−92.92	−86.25
butanolum	−99.79	−99.09	pimetine hydrochloride	−93.90	−98.24
cupriaseptol	−99.70	−99.36	itanoxone	−97.57	−92.78
etamsylate	−96.73	−97.71	fluperlapine	−90.64	−30.72
		Test Inactive Group			
dimaval	−99.96	−99.51	orciprenaline sulfate	−79.18	−66.72
quisqua lamine	−99.68	−98.20	isofenefrine	−93.73	−87.49
thiacetazone	−85.50	−91.77	foscarnet	−99.99	−99.99
clioquinol	−90.38	−76.41	betaxolol	−35.66	−5.06
stilonium iodide	−72.25	−71.00	betaquil	−99.68	−99.82
halothane	−99.39	−99.97	luvatren	−97.45	−91.76
cloral betaine	−99.85	−99.98	hydramitrazine	65.98	13.60
morantel tartrate	−98.67	−96.56	dioxifedrine	−92.10	−89.97

Table 6 (Continued)

name	$\Delta P\%^a$	$\Delta P\%^b$	name	$\Delta P\%^a$	$\Delta P\%^b$
furazolidone	80.62	38.98	exaprolol hydrochloride	−23.94	20.35
centpiperalone	−98.48	−92.33	mephentermine	−98.32	−98.90
clomoxir sodium	−86.48	−88.61	tetryzoline hydrochloride	−98.50	−96.42
clofibrate	−82.10	−59.78	apoatropine	−91.37	−90.15
ciproximide	−98.47	−92.69	fenoverine	−95.61	−44.62
aloe-emodin	−75.22	29.25	propanoic acid	−99.91	−99.87
oxyphenisatin Acetate	−75.54	−24.82	fluorembichin	−99.44	−96.30
oxaden	−75.55	−88.66	novembitol	−99.05	−97.06
chlormerodrin	−98.47	−99.61	nifuron	96.33	93.92
phensuximide	−99.67	−98.66	vinconate	−95.37	−74.79
dimepheptanol	−54.81	−76.41	canfochinid	−56.15	−61.41
clofenapic acid	−92.30	−91.87	diisopromine	−29.24	−71.82
vinamar	−99.80	−99.60	acetylcholine	−99.75	−99.90
saligenol	−99.72	−99.46	dithiophos	−69.65	14.35
percloroetane	−97.39	−99.89	hydroxyamfetamine	−99.01	−96.23
dimetrizadole	−54.09	−70.09	phenylephrine	−99.27	−98.91
guanoxabenz	−81.66	−97.83	tymazoline hydrochloride	−92.83	−77.07

[a] Results of the classification of compounds obtained from eq 11 (using nonstochastic quadratic indices). [b] Results of the classification of compounds obtained from eq 12 (using stochastic quadratic indices): $\Delta P\% = [P(\text{Active}) - P(\text{Inactive})]100$ (see developing classification functions).

Table 7. Intercorrelation of the Molecular Descriptors Included in the LDA-Based QSAR Models

nonorthogonal atom, atom-type, and total nonstochastic quadratic indices								
$q_0(x)$	$q_1(x)$	$q_{2L}(x_E)$	$q_0^H(x)$	$q_2^H(x)$	$q_{1L}^H(x_E)$	$q_{2L}^H(x_E)$	$q_{1L}(x_{E-H})$	$q_{15L}(x_{E-H})$
1.00	0.97	0.61	0.96	0.96	0.63	0.64	0.01	0.17
	1.00	0.57	0.91	0.98	0.59	0.60	0.00	0.25
		1.00	0.46	0.49	0.96	0.98	0.23	0.37
			1.00	0.97	0.50	0.51	−0.05	0.07
				1.00	0.52	0.53	−0.04	0.18
					1.00	0.98	0.37	0.42
						1.00	0.27	0.37
							1.00	0.61
								1.00

nonorthogonal atom, atom-type, and total stochastic quadratic indices									
${}^sq_{1L}(x_E)$	${}^sq_0^H(x)$	${}^sq_1^H(x)$	${}^sq_2^H(x)$	${}^sq_3^H(x)$	${}^sq_{1L}^H(x_E)$	${}^sq_0(x_{E-H})$	${}^sq_{1L}(x_{E-H})$	${}^sq_{2L}(x_{E-H})$	${}^sq_{3L}(x_{E-H})$
1.00	0.57	0.53	0.56	0.54	0.93	0.18	0.18	0.18	0.19
	1.00	0.96	0.96	0.94	0.46	−0.05	−0.05	−0.06	−0.06
		1.00	0.98	0.95	0.43	−0.05	−0.05	−0.06	−0.06
			1.00	0.97	0.45	−0.05	−0.06	−0.06	−0.06
				1.00	0.44	−0.05	−0.06	−0.06	−0.06
					1.00	0.49	0.49	0.49	0.50
						1.00	0.97	0.96	0.94
							1.00	0.98	0.95
								1.00	0.97
									1.00

the statistical interpretation of the models by using interrelated indices.[89−95] This process is an approach in which molecular descriptors are transformed in such a way that they do not mutually correlate. The main philosophy of this approach is to avoid the exclusion of descriptors on the basis of their colinearity with other variables previously included in the model. Both, the nonorthogonal descriptors and the derived orthogonal descriptors contain the same information. In this sense, the same statistical parameters of the QSAR models are obtained.[89−95] It is known that the interrelation among the different descriptors can result in highly unstable regression coefficients, which makes it impossible to know the relative importance of an index included in a model. However, in some cases, strongly interrelated descriptors can enhance the quality of a model because the small fraction of a descriptor that is not reproduced by its strongly interrelated pair can provide positive contributions to the modeling. On the other hand, the coefficient of the QSAR model based on orthogonal descriptors are stable to the inclusion of novel descriptors, which permits the interpretation of the regression coefficients and the evaluation of the role of individual fingerprints in the QSAR model.

In Table 8, we show the results of the orthogonalization of molecular descriptors included in both models. In this case, eqs 11a and 12a correspond to the final models with the orthogonalized molecular indices (see Table 8). Here, we used the symbols ${}^mO[q_k(x)]$, where the superscript m expresses the order of importance of the variable $[q_k(x)]$ after a preliminary forward stepwise analysis and O means orthogonal.

It must be highlighted here that the orthogonal descriptor-based models coincide with the collinear (i.e., ordinary) TOMOCOMD−CARDD descriptor-based models in all statistical parameters. The statistical coefficients of LDA QSARs λ, F, MCC, accuracy, %(+) (good classifications in the active group), and %(−) (good classifications in the inactive group) are the same whether we use a set of nonorthogonal descriptors or the corresponding set of

Table 8. Results of Randić's Orthogonalization Analysis

orthogonal atom, atom-type, and total nonstochastic quadratic indices

$^{1}O[q_0^{H}(x)]$	$^{2}O[q_2^{H}(x)]$	$^{3}O[q_{15L}(x_{E-H})]$	$^{4}O[q_{1L}(x_{E-H})]$	$^{5}O[q_1(x)]$	$^{6}O[q_0(x)]$	$^{7}O[q_{2L}^{H}(x_E)]$	$^{8}O[q_{2L}(x_E)]$	$^{9}O[q_{1L}^{H}(x_E)]$
1	0	0	0	0	0	0	0	0
	1	0	0	0	0	0	0	0
		1	0	0	0	0	0	0
			1	0	0	0	0	0
				1	0	0	0	0
					1	0	0	0
						1	0	0
							1	0
								1

LDA-based model derived with orthogonal atom, atom-type, and total nonstochastic quadratic indices

$$\textit{Class} = -0.15069 + 4.7535^{1}O[q_0^{H}(x)] - 3.80426^{2}O[q_2^{H}(x)] + 1.17955^{3}O[q_{15L}(x_{E-H})] - 2.36650^{4}O[q_{1L}(x_{E-H})] + 6.22277^{5}O[q_1(x)] - 15.73721^{6}O[q_0(x)] - 0.97037^{7}O[q_{2L}^{H}(x_E)] + 11.35210^{8}O[q_{2L}(x_E)] + 12.44961^{9}O[q_{1L}^{H}(x_E)] \quad \textbf{(11.a)}$$

$N = 1120$, $\lambda = 0.32$, $D^2 = 3.9$, $F = 258.32$, MCC = 0.89, Accuracy (%) = 94.73, %(+) = 96.34, %(−) = 92.22, $p < 0.0001$

orthogonal atom, atom-type, and total stochastic quadratic indices

$^{1}O[^{s}q_0^{H}(x)]$	$^{2}O[^{s}q_2^{H}(x)]$	$^{3}O[^{s}q_1^{H}(x)]$	$^{4}O[^{s}q_{1L}(x_{E-H})]$	$^{5}O[^{s}q_{1L}^{H}(x_E)]$	$^{6}O[^{s}q_{1L}(x_E)]$	$^{7}O[^{s}q_3^{H}(x_{E-H})]$	$^{8}O[^{s}q_0(x_{E-H})]$	$^{9}O[^{s}q_{3l}(x_{E-H})]$	$^{10}O[^{s}q_{2L}(x_{E-H})]$
1	0	0	0	0	0	0	0	0	0
	1	0	0	0	0	0	0	0	0
		1	0	0	0	0	0	0	0
			1	0	0	0	0	0	0
				1	0	0	0	0	0
					1	0	0	0	0
						1	0	0	0
							1	0	0
								1	0
									1

LDA-based model derived with orthogonal atom, atom-type, and total stochastic quadratic indices

$$\textit{Class} = -0.5589 + 3.9445^{1}O[^{s}q_0^{H}(x)] - 54.2074^{2}O[^{s}q_2^{H}(x)] + 40.3252^{3}O[^{s}q_1^{H}(x)] - 0.8304^{4}O[^{s}q_{1L}(x_{E-H})] + 1.7579^{5}O[^{s}q_{1L}^{H}(x_E)] - 4.7052^{6}O[^{s}q_{1L}(x_E)] - 34.7293^{7}O[^{s}q_3^{H}(x)] + 3.9482^{8}O[^{s}q_0(x_{E-H}] - 5.7690^{9}O[^{s}q_{3L}(x_{E-H}] - 8.4606^{10}O[^{s}q_{2L}(x_{E-H})] \quad \textbf{(12.a)}$$

$N = 1120$, $\lambda = 0.35$, $D^2 = 7.68$, $F = 203.11$, MCC = 0.86, Accuracy (%) = 93.13, %(+) = 90.16, %(−) = 95.02, $p < 0.0001$

orthogonal indices. This is not surprising because the latter models are derived as a combination of the former ones and cannot have more information content than them.[89−91] Only the D^2 values were different in both equation sets. This is because, before carrying out the orthogonalization process, all of the variables were standardized. In standardization, all values of the selected variables (molecular descriptors) were replaced by standardized values, which are computed as follows: standard score = (raw score − mean)/standard deviation. LDA algorithms at one point need to assess the distances between groups' centroids (or between cases and centroids), and obviously, when computing D^2 distances, LDA needs to decide on a scale. Because the different molecular fingerprints included here used entirely "different types of scales", the data were standardized so that each variable has a mean of 0 and a standard deviation of 1. This fact also makes possible the interpretation of the coefficients in the LDA-QSAR equations. Therefore, $^{m}O[q_k(x)]$ may be classified according to the distance k into short- (0−5), mid- (6−10), and long-range nonstochastic and stochastic quadratic indices. The information in Table 8 clearly shows that the major contribution to antimalarial activity is providing by short-range TOMOCOMD−CARDD descriptors.

4.3. Comparative Analysis of the Obtained Structure-Based Classification Models for Describing the Antimalarial Activity of a Heterogeneous Series of Compounds. In a previous paper, some of the present authors reported two classification models of antimalarial activity using the same training data set but including nonstochastic and stochastic linear indices.[27] With the aim to comparatively evaluate the ability of the nonstochastic and stochastic quadratic indices to encode chemical information and the quality of the obtained LDA-based classification models, we performed an examination of some statistical parameters. Table 9 summarizes the main results achieved with both TOMOCOMD−CARDD descriptors (based on both quadratic and linear maps).

Making use of the models obtained here (eqs 11 and 12), which includes nonstochastic and stochastic quadratic indices, 94.73% and 93.13% of the compounds in the training dataset were correctly classified. As can be observed in Table 9, models 13 and 14, obtained considering nonstochastic and stochastic linear indices,[27] show lower values for such parameters [accuracy of 94.02% (93.42%) and 91.52% (90.50%) in the training (test) set, respectively]. Also, the models reported in this work showed higher MCCs than the models obtained in our previous study. As can be seen, models developed with quadratic maps-based TOMOCOMD−CARDD descriptors (eqs 11 and 12) showed better parameters in all cases than models developed with linear maps-based TOMOCOMD−CARDD indices (eqs 13 and 14; see also eqs 10 and 11 in ref 27). In this sense, we can conclude that with the use of quadratic indices, it is possible to codify useful chemical information and obtain classification models comparable to or even better than those obtained using analogous descriptors already reported. Nevertheless, these differences are not very significant and only a future analysis of multiple data sets will indicate whether there is any advantage between them.

Table 9. Comparative Analysis of the Obtained Structure-Based Classification Models for Describing the Antimalarial Activity of a Heterogeneous Series of Compounds

models' features to be compared[a]	structure-based classification models of antimalarial activity					
	eq 11	eq 12	eq 13	eq 14	eq 15	eq 16
N total	1562	1562	1562	1562	59	60
N antimalarials	597	597	597	597	25	25
technique[b]	LDA	LDA	LDA	LDA	LDA	LDA
Wilks' λ (U statistics)	0.32	0.35	0.35	0.38	0.55	0.35
F	258.32	203.11	261.61	202.73	9.83	8.88
D^2	8.8	7.7	7.92	6.90		
p level	<0.0001	<0.0001	<0.0001	<0.0001		
			Training Set			
N total	1120	1120	1120	1120	41	45
N antimalarials	437	437	437	437	17	19
accuracy (%)	94.73	93.13	94.02	91.52	82.92	91.11
MCC[c]	0.89	0.86	0.87	0.82	0.65	0.82
families of drugs[d]	broader range	broader range	broader range	broader range	low range	low range
			Test Set			
N total	442	442	442	442	18	15
N antimalarials	160	160	160	160	8	6
predictability (%)	93.89	93.44	93.42	90.50	88.88	60.00
MCC[c]	0.87	0.86	0.86	0.79	0.92	0.22
families of drugs[d]	broader range	broader range	broader range	broader range	low range	low range

[a] Equations 11 and 12 are reported in this work, and models 13 and 14 were obtained previously by the present authors using nonstochastic and stochastic linear indices.[27] Equations 15 and 16 were reported by Gozalbez et al.[6] for two different studies: eq 15 was performed for the classification of antimalarial drugs and nonantiprotozoan drugs and eq 16 for the discrimination between antimalarials and antiprotozoan drugs without antimalarial activity. [b] LDA refers to linear discriminant analysis. [c] Matthews' correlation coefficient. [d] Only largely represented families were considered.

On the other hand, in the past decade, two other in silico methods have also been used to develop structure-based classification models (eqs 15 and 16 in Table 9) of antimalarial activity, which give rise to a good discrimination of this activity in large and heterogeneous series of organic compounds.[6] We also compare both approaches in order to show the potentialities of our method. In this case, because of differences in the composition of the experimental data used in carrying out the QSAR, it is not feasible to perform a "strict" comparison between the method reported previously[6] and the current approach. However, a relative comparison could be based on the kind of method used for deriving the QSAR and their statistical parameters, the number and diversity of chemical structural patterns contained in the data, the overall accuracy (%), the MCC, and the method that was used for the validation of the models. Table 9 also shows these chemometric coefficients for all approaches.

The global good classification in the training set of quadratic maps-based TOMOCOMD−CARDD models was higher than the two reported LDA equations (see Table 9). It is remarkable that the TOMOCOMD−CARDD models were derived from training series 27.3 (1120/41) and 24.8 (1120/45) times bigger than the series used by Gozalbes et al.[6] In this sense, the overall accuracy in test sets of quadratic maps-based TOMOCOMD−CARDD models was higher than the rest of two reported LDA equations (see Table 9).

Another remarkable aspect is the reference to the spectrum of structural patterns considered in the studies under comparison. Without a doubt, for the development of the TOMOCOMD−CARDD models reported here, a broader diversity of antimalarial compounds was considered.

4.4. Virtual Screening of Ras FTase Inhibitors: An Experiment of Lead Generation. One of the most important aspects of any QSAR model is its ability to predict the desired activity for new compounds not included in the training data set. Virtual screening of large databases considering the use of such models has emerged as an interesting alternative to high-throughput screening and an important drug-design tool.[10,11,102−104] With the aim of testing the ability of our models to detect new lead compounds with "unknown" structures, we carried out a simulated virtual screening of inhibitors of farnesyltransferase (FTase) that showed potent antimalarial activity in cell assays.[105] Not a single compound with this kind of structure was included in the training data set, and in this sense, this evaluation is equivalent to the discovery of new lead compounds using the developed models.

In this simulation, 10 previously reported FTase inhibitors with potent antimalarial activity were evaluated with models 11 and 12 as active/inactive ones. The results of the classification are shown in Table 10, and the molecular structures are illustrated in Chart 1.

As it can be seen, both models correctly classify most of the 10 selected compounds. In the first case, only three FTase inhibitors were classified as false inactives (70% of correct classification), whereas with model 12, the prediction has an overall accuracy of 100%.

This result is in accordance with the character of the TOMOCOMD−CARDD approach, which permits the consideration of implicitly, through the calculation of nonstochastic and stochastic quadratic molecular descriptors, substructural and global features responsible for a specific activity. In this way, new lead compounds could be designed using the TOMOCOMD−CARDD method described in this paper.

4.5. Experimental Results: Discovery of Novel Quinolinic Intermediaries as Antimalarial Compounds. The aim of the present work is the development of discriminant functions for the rational design (or selection/identification) of new antimalarial compounds via virtual screening. As shown, we explored the ability of our classification models

Chart 1. Molecular Structures of Peptidomimetic Inhibitors of Protein FTase that Showed Potent Antimalarial Activity in Cell Assays (See Table 10)

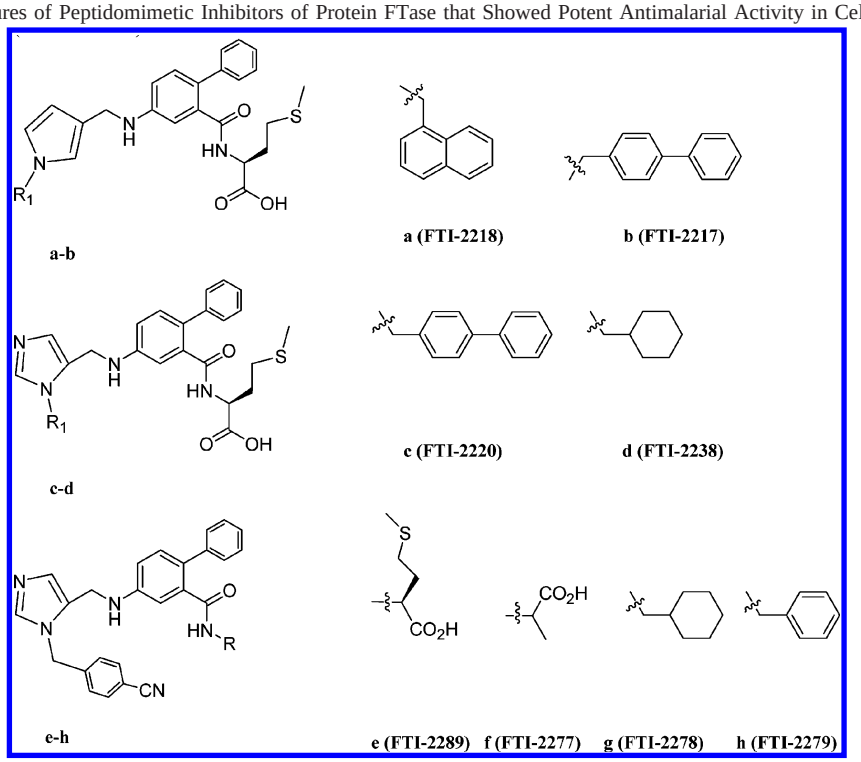


Table 10. Results of the Virtual Screening Simulation of Peptidomimetic Inhibitors of Protein FTase that Showed Potent Antimalarial Activity in Cell Assays

compound[a]	*P. falciparum*-infected RBC ED_{50} (μg/mL)[b]	eq 11		eq 12	
		(ΔP%)[c]	class	(ΔP%)[d]	class
a (FTI-2218)	10	88.40	+	88.05	+
b (FTI-2217)	12	93.44	+	90.88	+
c (FTI-2220)	5	97.13	+	92.63	+
d (FTI-2238)	10	27.23	+	74.48	+
e (FTI-2289)	13	−39.22	−	19.46	+
f (FTI-2277)	3	69.39	+	90.44	+
g (FTI-2278)	3	−25.08	−	26.14	+
h (FTI-2279)	4	−61.68	−	38.75	+
i (FTI-2291)	10	18.90	+	36.45	+
j (FTI-2153)	2	21.28	+	58.45	+

[a] Compounds a−j were taken from Ohkanda et al.[105] [b] Inhibition at 20 μM; RBC = red blood cell. [c] Results of the classification of compounds obtained from eq 11. [d] Results of the classification of compounds obtained from eq 12.

to find new active compounds carrying out an experiment of lead generation for the case of Ras FTPase inhibitors. These results encouraged us to develop a search for novel active compounds not described yet as antimalarials in the literature.

In this sense, we also explore a large dataset of organic chemicals through virtual screening in order to discover novel candidates for antimalarial drug-like compounds. A great number of the candidates to be assayed as antimalarial, detected with our models, were sent to biological assays, and their presentation will be the objective of a forthcoming paper. Nevertheless, in this work, we want to show some promissory outcomes of this computational screening, which can represent an important starting point to the design of novel antimalarials.

It is well-known that the majority of compounds used in the treatment of malaria are quinolinic derivatives such as quinine, chloroquine, mefloquine, halofantrine, and primaquine. Acyclic β-enaminoesters and arylaminomethylenemalonates are synthetic intermediates of quinolinic compounds and can be achieved by economic and simple synthetic routes.[106,107] On the other hand, there is not much research related to the biological activity of enamine compounds. Taking that into account, we explored in our search the behavior of some acyclic β-enamino esters and arylaminomethylenemalonates. Three of these compounds were initially evaluated with models 11 and 12 and, in order to corroborate the predictions, prepared with excellent yields by very economic and simple methods and evaluated against two strains of *P. falciparum.*

The acyclic β-enamino ester **1** was prepared by means of a nucleophilic addition of the aromatic amine to the keto group of the corresponding β-keto ester, using a previously described methodology.[108] Arylaminomethylenemalonates were synthesized by means of a "one pot" process, starting from equimolar quantities of the corresponding aniline, ethyl malonate, and ethyl orthoformate in the presence of a catalytic amount of $ZnCl_2$.[109] Both general procedures are shown in Scheme 1. All of the structures were confirmed by spectroscopic data analysis, which is given as Supporting Information.

The results of the prediction process using models 11 and 12, as well as the MIC for the three assayed compounds against K1 and Palo Alto strains, are shown in Table 11.

The sensitivity control of each strain was carried out with chloroquine diphosphate. The MIC of chloroquine for sensitive strains is 5.7 pmol/well; that is, strains with a MIC above of this value are resistant to this compound.[110] In our study, the determined value of the MIC for the K1 strain was 8 pmol/well (μmol/L) and that for the Palo Alto strain was 4 pmol/well (0.8 μmol/L), confirming the sensitivity of the used strains.

Scheme 1. Synthetic Procedure for the Synthesis of Quinolinic Intermediaries

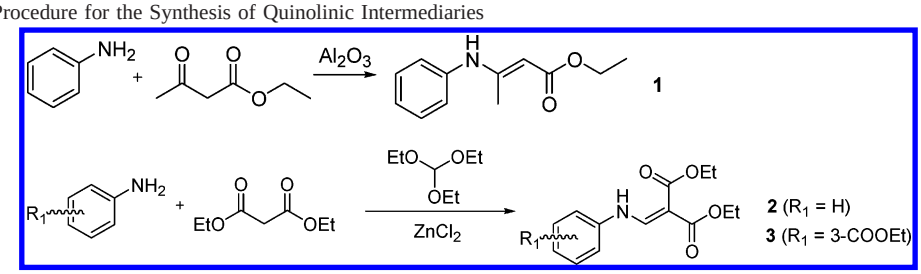
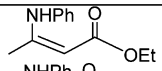
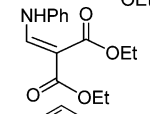
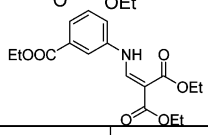
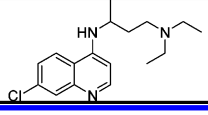


Table 11. Synthetic Intermediates of Quinolinic Compounds Evaluated in the Present Study, Their Classifications (Δ*P*%) According to the TOMOCOMD-CARDD Approach, Their Antimalarial Activities against Two Malarial Strains, and the Antimalarial Activity of Chloroquine

Compound	Structure	ΔP%[a]	class	ΔP%[b]	class	MIC(pmol/well)	
						K1	Palo Alto
1		-93.09	-	-78.83	-	100	100
2		-63.02	-	-1.25	-	>100	>100
3		**31.21**	+	**89.02**	+	**32**	**16**
Chloroquine		**77.93**	+	**77.96**	+	**8**	**4**

[a] Results of the classification of compounds obtained from eq 11. [b] Results of the classification of compounds obtained from eq 12.

As expected, compound **1** did not show activity against the K1 and Palo Alto strains. The inhibition of the schizont maturation was observed at 100 pmol/well. Compound **2** did not inhibit the growth of parasites at any of the assayed concentrations (MIC > 100 pmol/well). Conversely, and in accordance with the predictions, the best results were observed for compound **3**, which showed a MIC = 32 pmol/well against the K1 strain and a MIC = 16 pmol/well for the case of the Palo Alto strain.

Taking into account that this is the first report of an arylaminomethylenemalonate with antimalarial activity, the result can be considered as a very promising starting point for the future design and refinement of novel compounds with higher antimalarial activity. That is to say, compound **3** was tested at higher doses than chloroquine diphosphate (reference or control antimalarial drug), but this result leaves a door open to a virtual variation study of the structure of these compounds in order to improve their antimalarial activity. Other chemicals in the same family as compound **3**, as well as other chemicals not in this family, were also predicted as antimalarials. The synthesis, characterization, and biological evaluation of these compounds are, however, beyond the scope of the present paper and will be discussed elsewhere. It is important to recall that the aim of this study is not to validate the model but to provide an experimental example of how to use the model for potential drug discovery.

5. CONCLUDING REMARKS

The introduction and use of graph theoretical descriptors for rational drug design has become an attractive tool for medicinal chemists. In this sense, the fusion of high throughput screening and classification-based QSAR models in an attempt to minimize the costs in terms of time, financial, human, and animal resources is becoming a viable alternative to massive screening. In this work, we have shown that the TOMOCOMD−CARDD approach can be applied to generate useful quantitative models for the classification of antimalarials. In a flexible way, this method permits a quick in silico discovery of new candidates as possible lead compounds making use of a minimum of resources. Considering a training data set of compounds with a considerable structural variability, we reduce the degree of uncertainly for this process. The simulated virtual screening of Ras FTase inhibitors with antimalarial properties has proved the ability of our models for an adequate discrimination of new active compounds from inactive ones. The collected data of the active compounds used in this study results in an important tool not only for the theoretical research but for the general scientific work in this area.

Using the developed models, we have identified a new lead candidateas a promising starting point for the design of new arylaminomethylenemalonates with potent antimalarial activity. Some work in this direction is currently in progress and will be published in a forthcoming paper.

The interactive character of the TOMOCOMD−CARDD approach permits the future inclusion of new antimalarial drugs in the training data set and the generation ofmore "intelligent" models each time. In this sense, the new considered structural patterns will be recognized for the models and a better discrimination of such kinds of compounds will be obtained. However, this point is out of the scope of the present work.

Supporting Information Available: The complete list of compounds used in the training and prediction sets, as well as their structures, posterior classification according to models 11 and 12, chemistry, and data analysis of the obtained chemicals is available free of charge via the Internet at http://pubs.acs.org.

ACKNOWLEDGMENT

Thanks are given to the FAPESP, CAPES, and CNPq for partial financial support. The authors sincerely acknowledge the kind attention of the editor, Prof. Dr. Anton J. Hopfinger, and the useful comments from unknown reviewers, which contributed to an increase in the quality of the present paper.

REFERENCES AND NOTES

(1) Walsh, J. A. Disease Problems in the World. *Ann. N.Y. Acad. Sci.* **1989**, *569*, 1−16.

(2) Torok, D. S.; Ziffer, H. Synthesis and Antimalarial Activities of N-Substituded 11-Azaartemisinins. *J. Med. Chem.* **1995**, *38*, 5045−5050.

(3) Posner, G. H.; O'Dowd, H.; Ploypradith, P.; Cumming, J. N.; Xie, S.; Shapiro, T. A. Antimalarial Cyclic Peroxy Ketals. *J. Med. Chem.* **1998**, *41*, 2164−2167.

(4) Posner, G. H.; Cumming, J. N.; Woo, S. H.; Ploypradith, P.; Xie, S.; Shapiro, T. A. Orally Active Antimalarial 3-Substituted Trioxanes: New Synthetic Methodology and Biological Evaluation. *J. Med. Chem.* **1998**, *41*, 940−951.

(5) Lin, A. J.; Zikry, A. B.; Kyle, D. E. Antimalarial Activity of New Dihydroartemisinin Derivatives. 7. 4- (p-Substituted phenyl)-4 (R or S)-[10 (alpha or beta)-hydroartemisininoxy]butyric Acids. *J. Med. Chem.* **1997**, *40*, 1396−1400.

(6) Gonzalbes, R.; Gálvez, J.; Moreno, A.; García-Domenech, R. Discovery of New Antimalarial Compounds by Use of Molecular Connectivity Techniques. *J. Pharm. Pharmacol.* **1999**, *52*, 111−117.

(7) Go, M. L.; Ngiam, T. L.; Tan, A. L. C.; Kuaha, K.; Wilairat, P. Structure−Activity Relationships of some indolo(3,2-c)quinolines with Antimalarial Activity. *Eur. J. Pharm. Sci.* **1998**, *6*, 19−26.

(8) McKie, J. H.; Douglas, K. T.; Chan, C.; Roser, S. A.; Yates, R.; Read, M.; Hyde, J. E.; Dascombe, M. J.; Yuthavong, Y.; Sirawaraporn, W. Rational Drug Desing Approach for Overcoming Drug Resistance: Application to Pyrimethamine Resistance in Malaria. *J. Med. Chem.* **1998**, *41*, 1367−1370.

(9) De, D.; Krogstad, F. M.; Byers, L. D.; Krogstad, D. J. Structure−Activity Relationships for Antiplasmodial Activity Among 7-Substituted 4- Aminoquinolines. *J. Med. Chem.* **1998**, *41*, 4918−4926.

(10) Seifert, M. H. J.; Wolf, K.; Vitt, D. Virtual High-Throughput *in silico* Screening. *BIOSILICO* **2003**, *I*, 143−149.

(11) Xu, J.; Hagler, A. Chemoinformatics and Drug Discovery. *Molecules* **2002**, *7*, 566−600.

(12) Estrada, E.; Uriarte, E.; Montero, A.; Teijeira, M.; Santana, L.; De Clercq, E. A Novel Approach for Virtual Screening and Rational Design of Anticancer Compounds. *J. Med. Chem*. **2000**, *43*, 1975−1985**.**

(13) González-Díaz, H.; Marrero-Ponce, Y.; Hernández, I.; Bastida, I.; Tenorio, E.; Nasco, O.; Uriarte, U.; Castañedo, N.; Cabrera, M. A.; Aguila, E.; Marrero, O.; Morales, A.; Pérez, M. 3D-MEDNEs: An Alternative "In Silico" Technique for Chemical Research in Toxicology. 1. Prediction of Chemically Induced Agranulocytosis. *Chem. Res. Toxicol.* **2003**, *16*, 1318−1327.

(14) de Julián-Ortiz, J. V.; de Alapont, C. G.; Ríos-Santamarina, I.; García-Doménech, R.; Gálvez, J. Prediction of Properties of Chiral Compounds by Molecular Topology. *J. Mol. Graphics Modell.* **1998**, *16*, 14−18.

(15) Marrero-Ponce, Y.; Romero, V. *TOMOCOMD*; Central University of Las Villas: Villa Clara, Cuba, 2002. TOMOCOMD (***TO***pological ***MO***lecular ***COM***puter ***D***esign) for Windows, version 1.0 is a preliminary experimental version; in the future, a professional version can be obtained upon request to Y. Marrero: yovanimp@qf.uclv.edu.cu or ymarrero77@yahoo.es.

(16) Marrero-Ponce, Y. Total and Local Quadratic Indices of the Molecular Pseudograph's Atom Adjacency Matrix: Applications to the Prediction of Physical Properties of Organic Compounds. *Molecules* **2003**, *8*, 687−726.

(17) Marrero-Ponce, Y. Linear Indices of the "Molecular Pseudograph's Atom Adjacency Matrix": Definition, Significance-Interpretation and Application to QSAR Analysis of Flavone Derivatives as HIV-1 Integrase Inhibitors. *J. Chem. Inf. Comput. Sci*. **2004**, *44*, 2010−2026.

(18) Marrero-Ponce, Y. Total and Local (Atom and Atom-Type) Molecular Quadratic Indices: Significance-Interpretation, Comparison to Other Molecular Descriptors and QSPR/QSAR Applications. *Bioorg. Med. Chem*. **2004**, *12*, 6351−6369.

(19) Marrero-Ponce, Y.; Castillo-Garit, J. A.; Torrens, F.; Romero-Zaldivar, V.; Castro, E. Atom, Atom-Type and Total Linear Indices of the "Molecular Pseudograph's Atom Adjacency Matrix": Application to QSPR/QSAR Studies of Organic Compounds. *Molecules* **2004**, *9*, 1100−1123.

(20) 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 σ-Receptor Antagonist Activities. *Bioorg. Med. Chem.* **2004**, 12, 5331−5342.

(21) Marrero-Ponce, Y.; Cabrera, M., A.; Romero, V.; Ofori, E.; Montero, L. A. Total and Local Quadratic Indices of the "Molecular Pseudograph's Atom Adjacency Matrix". Application to Prediction of Caco-2 Permeability of Drugs. *Int. J. Mol. Sci*. **2003**, *4*, 512−536.

(22) Marrero-Ponce, Y.; Cabrera, M. A.; Romero, V.; González, D. H.; Torrens, F. A New Topological Descriptors Based Model for Predicting Intestinal Epithelial Transport of Drugs in Caco-2 Cell Culture. *J. Pharm. Pharm. Sci.* **2004**, *7*, 186−199.

(23) 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 Electronics J. Mol. Des.* **2005**, *4*, 124−150.

(24) 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*. **2005**, *13*, 1005−1020.

(25) 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. Des.* **2004**, *18*, 615−633.

(26) 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. *THEOCHEM* **2005**, *717*, 67−79.

(27) Marrero-Ponce, Y.; Montero-Torres, A.; Romero-Zaldivar, C.; Iyarreta-Veitía, I.; Mayón Peréz, 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.* **2005**, *13*, 1293−1304.

(28) Marrero-Ponce, Y.; Medina-Marrero, R.; Torrens, F.; Martinez, Y.; Romero-Zaldivar, V.; Castro, E. A. Atom, Atom-type, and Total Non-Stochastic and Stochastic Quadratic Fingerprints: A Promising Approach for Modeling of Antibacterial Activity. *Bioorg. Med. Chem*. **2005**, *13*, 2881−2899.

(29) 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. Modell.* Accepted for publication.

(30) 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 Ψ-RNA Packaging Region. *Int. J. Mol. Sci.* **2004**, *5*, 276−293.

(31) 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 Ψ-RNA Packaging Region. *Bioorg. Med. Chem.* **2005**, *13*, 3397−3404.

(32) 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 α-Carbon Atom Adjacency Matrix". 1. Prediction of Arc Repressor Alanine-mutant's Stability. *Molecules* **2004**, *9*, 1124−1147.

(33) 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 α-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.* **2005**, *13*, 3003−3015.
(34) Pauling, L. *The Nature of Chemical Bond*; Cornell University Press: New York, 1939; pp 2−60.
(35) Walker, P. D.; Mezey, P. G. Molecular Electron Density Lego Approach to Molecule Building. *J. Am. Chem. Soc.* **1993**, *115*, 12423−12430.
(36) Golbraikh, A.; Bonchev, D.; Tropsha, A. Novel Chirality Descriptors Derived from Molecular Topology. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 147−158.
(37) González-Díaz, H.; Hernández-Sánchez, I.; Uriarte, E.; Santana, L. Symmetry Considerations in Markovian Chemicals 'in silico' Design (MARCH−INSIDE) I.: Central Chirality Codification, Classification of ACE Inhibitors and Prediction of σ-Receptor Antagonist Activities. *Comput. Biol. Chem.* **2003**, *27*, 217−227.
(38) Klein, D. J. Graph Theoretically Formulated Electronic-Structure Theory. *Internet Electronics J. Mol. Des*. **2003**, *2*, 814−834; http://www.biochempress.com.
(39) González-Díaz, H.; Bastida, I.; Castañedo, N.; Nasco, O.; Olazabal, E.; Morales, A.; Serrano, H. S.; Ramos, de A. R. Simple Stochastic Fingerprints Towards Mathematical Modelling in Biology and Medicine. 1. The Treatment of Coccidiosis. *Bull. Math. Biol.* **2004**, *66*, 1285−1311.
(40) Ann, D. C. In *Burger's Medicinal Chemistry and Drug Discovery*, 5th ed.; Wiley-Interscience Publication: New York, 1997; Vol. 5, Antimalarial Agents.
(41) Negwer, M. *Organic-Chemical Drugs and their Synonyms*; Akademie-Verlag: Berlin, 1987.
(42) Domínguez, J. N.; López, S.; Charris, J.; Iarroso, L.; Lobo, G.; Semenov, A.; Olson, J. E.; Rosenthal, P. J. Synthesis and Antimalarial Effects of Phenothiazine Inhibitors of a *Plasmodium falciparum* Cystein Protease. *J. Med. Chem.* **1997**, *40*, 2726−2732.
(43) Hawley, S. R.; Bray, P. G.; Mungthin, M.; Atkinson, J. D.; O'Neill, P. M.; Ward, S. A. Relationship Between Antimalarial Drug Activity, Accumulation, and Inhibition of Heme Polymerization in *Plasmodium falciparum* In Vitro. *Antimicrob. Agents Chemother.* **1998**, *42*, 682−686.
(44) Rucker, G.; Schenkel, E. P.; Manns, D.; Mayer, R.; Heiden, K.; Heinzmann, B. M. Sesquiterpene Peroxides from *Senecio selloi* and *Eupatorium rufescens*. *Planta Med*. **1996**, *62*, 565−566.
(45) Ring, C. S.; Sun, E.; Mekerrow, J. H.; Lee, G. K.; Rosenthal, P. J.; Kurtz, I. D.; Cohen, F. E. Structure-Based Inhibitor Design by Using Protein Models for the Development of Antiparasitic Agents. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 3583−3587.
(46) Ryley, J. F.; Peters, W. The Antimalarial of some Quinolone Esters. *Ann. Trop. Med. Parasitol.* **1970**, *64*, 209−222.
(47) Basco, L. K.; Dechy-Cabaret, O.; Ndounga, M.; Meche, F. S.; Robert, A.; Meunier, B. In vitro Activities of DU-1102, a New Trioxaquine DerivativeAgainst *Plasmodium falciparum* Isolates. *Antimicrob. Agents Chemother.* **2001**, *45*, 1886−1888.
(48) Haque, T. S.; Skillman, A. G.; Lee, C. E.; Habashita, H.; Gluzman, I. Y.; Swing, T. J. A.; Goldberg, D. E.; Knuts, I. D.; Ellman, J. A. Potent, Low-Molecular-Weight Non-peptide Inhibitors of Malarial Aspartyl Protease Plasmepsin II. *J. Med. Chem*. **1999**, *42*, 1428−1440.
(49) Raynes, K. Bisquinoline Antimalarials: Their Role in Malaria Chemotherapy. *Int. J. Parasitol.* **1999**, *29*, 367−379.
(50) Tsai, C. S.; Shen, A. Y. Synthesis and Bilogical Evaluation of Some Potential Antimalarials. *Arch. Pharm*. **1994**, *327*, 677−679.
(51) Posner, G. H.; Tao, X.; Cumming, J. N.; Klinedinst, D.; Shapiro, T. A. Antimalarially Potent, Easily Prepared, Fluorinated Endoperoxides. *Tetrahedron Lett.* **1996**, *37*, 7225−7228.
(52) Philipp, A.; Kepler, J. A.; Johnson, B. H.; Carroll, F. I. Peptide Derivatives of Primaquine as Potential Antimalarial Agents. *J. Med. Chem*. **1998**, *31*, 870−874.
(53) Figgitt, D.; Denny, W.; Chvalitshewinkoon, P.; Wilairat, P.; Ralph, R. In vitro Study of Anticancer Acridines as Potential Antitrypanosomal and Antimalarial Agents. *Antimicrob. Agents Chemother*. **1992**, *36*, 1644−1647.
(54) Nga, T. T. T.; Menage, C.; Begue, J. P.; Delpon, D. B.; Gantier, J. C. Synthesis and Antimalarial Activities of Fluoroalkyl Derivatives of Dihydroartemisinin. *J. Med. Chem*. **1998**, *41*, 4101−4108.
(55) Avery, M. A.; Bonk, J. D.; Chong, W. K. M.; Mehrota, S.; Miller, R.; Milhous, W.; Goins, D. K.; Venkatesan, S.; Wyandt, C.; Khan, I.; Avery, B. A. Structure−Activity Relationships of the Antimalarial Agent Artemisinin. 2. Effect of Heteroatom Substitution at O-11: Synthesis and Bioassay of N−Alkyl-11-aza-9-desmethylartemisinins. *J. Med. Chem*. **1995**, *38*, 5038−5044.
(56) Posner, G. H.; Wang, D.; González, L.; Tao, X.; Cumming, J. N.; Klinedints, D.; Shapiro, T. A. Mechanism-Based Design of Simple, Symmetrical, Easily Prepared, Potent Antimalarial Endoperoxides. *Tetrahedron Lett.* **1996**, *37*, 815−818.
(57) Pu, Y. M.; Torok, D. S.; Ziffer, H.; Pan, X. Q.; Meshnick, S. R. Synthesis and Antimalarial Activities of Several Fluorinated Artemisinin Derivatives. *J. Med. Chem*. **1995**, *38*, 4120−4124.
(58) Posner, G. H.; O'Dowd, H.; Caferro, T.; Cumming, J. N.; Ploypradith, P.; Xie, S.; Shapiro, T. A. Antimalarial Sulfone Trioxanes. *Tetrahedron Lett.* **1998**, *39*, 2273−2276.
(59) Posner, G. H.; McGarvey, D. J.; Oh, C. H.; Kumar, N.; Meshnick, S. R.; Asawamahasadka, W. Structure−Activity Relationships of Lactone Ring-Opened Analogues of the Antimalarial 1,2,4-Trioxane Artemisinin. *J. Med. Chem*. **1995**, *38*, 607−612.
(60) Posner, G. H.; González, L.; Cumming, J. N.; Klinedints, D.; Shapiro, T. A. Synthesis and Antimalarial Activity of Heteroatom-Containing Bicyclic Endoperoxides. *Tetrahedron* **1997**, *53*, 37−50.
(61) Venugopalan, B.; Bapat, C. P.; Karnik, P. J. Synthesis of A Nonel Ring Contracted Artemisinin Derivative. *Bioorg. Med. Chem. Lett.* **1994**, *4*, 751−752.
(62) Avery, M. A.; Gao, F.; Chong, W. K. M.; Hendrickson, T. F.; Inman, W. D.; Crews, P. Synthesis, Conformational Analysis, and Antimalarial Activity of Tricyclic Analogs of Artemisinin. *Tetrahedron* **1994**, *50*, 957−972.
(63) Avery, M. A.; Mehrotra, S.; Johnson, T. L.; Bonk, J. D.; Vroman, J. A.; Miller, R. Structure−Activity Relationships of the Antimalarial Agent Artemisinin. 5. Analogs of 10-Deoxoartemisinin Substituted at C-3 and C-9. *J. Med. Chem*. **1996**, *39*, 4149−4155.
(64) Venugopalan, B.; Bapat, C. P.; Karnik, P. J.; Chatterjee, D. K.; Iyer, N.; Lepcha, D. Antimalarial Activity of Novel Ring-Contracted Artemisinin Derivatives. *J. Med. Chem*. **1995**, *38*, 1992−1927.
(65) Zouhiri, F.; Desmaele, D.; d'Angelo, J.; Riche, C.; Gay, F.; Cicéron, L. Artemisinin Tricyclic Analogs: Role of a Methyl Group at C-5a. *Tetrahedron Lett.* **1998**, *39*, 2969−2972.
(66) Posner, G. H.; Parker, M. H.; Northrop, J.; Elias, J. S.; Ploypradith, P.; Xie, S.; Shapiro, T. A. Orally Active, Hydrolytically Stable, Semisynthetic, Antimalarial Trioxanes in The Artemisinin Family. *J. Med. Chem.* **1999**, *42*, 300−304.
(67) Cumming, J. N.; Wang, D.; Park, S. B.; Shapiro, T. A.; Posner, G. H. Design, Synthesis, Derivatization, and Structure−Activity Relationships of Simplified, Tricyclic, 1,2,4-Trioxane Alcohol Analogues of the Antimalarial Artemisinin. *J. Med. Chem*. **1998**, *41*, 952−964.
(68) Posner, G. H.; Oh, C. H.; Gerena, L.; Milhous, W. K. Extraordinarily Potent Antimalarial Compounds: New, Structurally Simple, easily Synthesized, Tricyclic 1,2,4-Trioxanes. *J. Med. Chem*. **1992**, *35*, 2459−2467.
(69) Calas, M.; Cordina, G.; Bompart, J.; Bari, M. B.; Jei, T.; Ancelin, M. L.; Vial, H. Antimalarial Activity of Molecules Interfering with *Plasmodium falciparum* Phospholipid Metabolism. Structure−Activity Relationships Analysis. *J. Med. Chem.* **1997**, *40*, 3557−3566.
(70) Ismail, F. M. D.; Dascombe, M. J.; Carr, P.; North, S. E. An Exploration of the Structure−activity Relationships of 4-Aminoquinolines: Novel Antimalarials with Activity. *J. Pharm. Pharmacol*. **1996**, *48*, 841−850.
(71) Ram, V. J.; Saxena, A. S.; Srivastavab, S.; Chandrab, S. Oxygenated Chalcones and Bischalcones as Potential Antimalarial Agents. *Bioorg. Med. Chem. Lett.* **2000**, *10*, 2159−2161.
(72) Posner, G. H.; Northrop, J.; Paik, I. H.; Borstnik, K.; Dolan, P.; Kensler, T. W.; Xiec, S.; Shapiroc, T. A. New Chemical and Biological Aspects of Artemisinin-Derived Trioxane Dimers. *Bioorg. Med. Chem.* **2000**, *10*, 227−232.
(73) Gironés, X.; Gallegos, A.; Carbó-Dorca, R. Antimalarial Activity of Synthetic 1,2,4-Trioxanes and Cyclic Peroxy Ketals, a Quantum Similarity Study. *J. Comput.-Aided Mol. Des.* **2001**, *15*, 1053−1063.
(74) Cheng, F.; Shen, J.; Luo, X.; Zhu, W.; Gu, J.; Ji, R.; Jiang, H.; Chen, K. Molecular Docking and 3-D-QSAR Studies on the Possible Antimalarial Mechanism of Artemisinin Analogues. *Bioorg. Med. Chem.* **2002**, *10*, 2883−2891.
(75) Santos-Filho, O. A.; Hopfinger, A. J. A Search for Sources of Drug Resistance by the 4D-QSAR Analysis of a Set of Antimalarial Dihydrofolate Reductase Inhibitors. *J. Comput.-Aided Mol. Des.* **2001**, *15*, 1−12.
(76) Jain, R.; Vangapandu, S.; Jain, M.; Kaur, N.; Singhb, S.; Singhb, P. P. Antimalarial Activities of Ring-Substituted Bioimidazoles. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 1701−1704.
(77) Itoh, T.; Shirakami, S.; Ishida, N.; Yamashita, Y.; Yoshida, T.; Kimb, H. S.; Watayab, Y. Synthesis of Novel Ferrocenyl Sugars and their Antimalarial Activities. *Bioorg. Med. Chem. Lett.* **2000**, *10*, 1657−1659.
(78) Reichenberg, A.; Wiesner, J.; Weidemeyer, C.; Dreiseidler, E.; Sanderbrand, S.; Altincicek, B.; Beck, E.; Schlitzerc, M.; Jomaa, H. Diaryl Ester Prodrugs of FR900098 with Improved In Vivo Antimalarial Activity. *Bioorg. Med. Chem. Lett.* **2001**, *11*, 833−835.
(79) Ryckebusch, A.; Deprez-Poulain, R.; Maes, L.; Debreu-Fontaine, M. A.; Mouray, E.; Grellier, P.; Sergheraert, C. Synthesis and in Vitro

and in Vivo Antimalarial Activity of N1-(7-Chloro-4-quinolyl)-1,4-bis(3-aminopropyl)piperazine Derivatives. *J. Med. Chem.* **2003**, *46*, 542−557.
(80) Nöteberg, D.; Hamelink, E.; Hulten, J.; Wahlgren, M.; Vrang, L.; Samuelsson, B.; Hallberg, A. Design and Synthesis of Plasmepsin I and Plasmepsin II Inhibitors with Activity in *Plasmodium falciparum*-Infected Cultured Human Erythrocytes. *J. Med. Chem.* **2003**, *46*, 734−746.
(81) Murray, P. J.; Kranz, M.; Ladlow, M.; Taylor, S.; Berst, F.; Holmes, A. B.; Keavey, K. N.; Laxa-chamiec, A.; Seale, P. W.; Stead, P.; Upton, R. J.; Croft, S. L.; Clegg, W.; Elsegood, M. R. J. The Synthesis of Cyclic Tetrapeptoid Analogues of the Antiprotozoal Natural Product Apicidin. *Bioorg. Med. Chem. Lett.* **2001**, *11*, 773−776.
(82) *The Merck Index*, twelfth ed.; Chapman & Hall: New York, 1996.
(83) McFarland, J. W.; Gans, D. J. In *Chemometric Methods in Molecular Design*; van Waterbeemd, H., Ed.; VCH Publishers: New York, 1995; pp 295−307, Cluster Significance Analysis.
(84) Johnson, R. A.; Wichern, D. W. *Applied Multivariate Statistical Analysis*; Prentice-Hall: New York, 1988.
(85) *STATISTICA*, version 5.5; Statsoft Inc.: Tulsa, OK, 1999.
(86) Wold, S.; Erikson, L. In *Chemometric Methods in Molecular Design*; van de Waterbeemd, H., Ed.; VCH Publishers: New York, 1995; pp 309−318, Statistical Validation of QSAR Results Validation Tools.
(87) Golbraikh, A.; Tropsha, A. Predictive QSAR Modelling Based on Diversity Sampling of Experimental Datasets for the Training and Test Set Selection. *J. Mol. Graphics Modell.* **2002**, *20*, 269−276.
(88) Baldi, P.; Brunak, S.; Chauvin, Y.; Andersen, C. A.; Nielsen, H. Assessing the Accuracy of Prediction Algorithms for Classification: an Overview. *Bioinformatics* **2000**, *16*, 412−424.
(89) Randić, M. Resolution of Ambiguities in Structure−Property Studies by Us of Orthogonal Descriptors. *J. Chem. Inf. Comput. Sci*. **1991**, *31*, 311−320.
(90) Randić, M. Orthogonal Molecular Descriptors. *New J. Chem*. **1991**, *15*, 517−525.
(91) Randić, M. Correlation of Enthalpy of Octanes with Orthogonal Connectivity Indices. *THEOCHEM* **1991**, *233*, 45−59.
(92) Lučić, B.; Nikolić, S.; Trinajstić, N.; Jurić, D. The Structure−Property Models can be Improved Using the Orthogonalized Descriptors. *J. Chem. Inf. Comput. Sci*. **1995**, *35*, 532−538.
(93) Klein, D. J.; Randić, M.; Babić, D.; Lučić, B.; Nikolić, S.; Trinajstić, N. Hierarchical Orthogonalization of Descriptors. *Int. J. Quantum Chem.* **1997**, *63*, 215−222.
(94) 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.* **2002**, *42*, 1194−1203.
(95) Estrada, E.; Uriarte, E. Recent Advances on the Role of Topological Indices in Drug Discovery Research. *Curr. Med. Chem.* **2001**, *8*, 1573−1588.
(96) Antonioletti, R.; Bonadies, F.; Orelli, L. O.; Scettri, A. Selective *C*-Alkylation of 1,3-dicarbonyl Compounds. *Gazz. Chim. Ital*. **1992**, *122*, 237−238.
(97) Perrin, D. D.; Armarego, W. L. F.; Perrin, D. R. *Purification of Laboratory Chemicals*; Pergamon Press: New York, 1980.
(98) Rieckmann, K. H.; Campbell, G. H.; Sax, L. J.; Mrema, J. E. Drug Sensitivity of *Plasmodium falciparum*. An in vitro Microtechnique. *Lancet* **1978**, *1*, 22−23.
(99) Trager, W.; Jensen, J. B. Human Malaria in Continuos Culture. *Science* **1976**, *193*, 673−675.
(100) Diggs, C.; Joseph, K.; Flemmings, B.; Snodgrass, R.; Hines, F. Protein Synthesis in Vitro by Cryopreserved *Plasmodium falciparium*. *Am. J. Trop. Med. Hyg.* **1975**, *24*, 760−763.
(101) Lambros, C.; Vanderberg, J. P. Synchronization of *Plasmodium falciparum* Erythrocytic Stages in Culture. *J. Parasitol*. **1979**, *65*, 418−420.
(102) Julián-Ortiz, J. V.; Gálvez, J.; Muñoz-Collado, C.; García-Domenech, R.; Gimeneo-Cardona, C. Virtual Combinatorial Synthesis and Computational Screening of New Potential Anti-Herpes Compounds. *J. Med. Chem.* **1999**, *42*, 3308−3314.
(103) Drie, J. H. V.; Lajiness, M. S. Approaches to Virtual Library Design. *Drug Discuss. Today.* **1998**, *3*, 274−283.
(104) Lajiness, M. In *Rouvray Computacional Chemical Graph Theory*; Rouvray, D. H., Ed.; Nova Science: New York, 1990; Molecular Similarity-Based Methods for Selecting Compounds for Screening.
(105) Ohkanda, J.; Lockman, J. W.; Yokoyama, K.; Gelb, M. H.; Croft, S. L.; Kendrich, H.; Harrell, M. I.; Feagin, J. E.; Blaskovich, M. A.; Sebti, S. M.; Hamilton, A. D. Peptidomimetic Inhibitors of Protein Farnesyltransferase Show Potent Antimalarial Activity. *Bioorg. Med. Chem. Lett.* **2001**, *11*, 761−764.
(106) Marvel, C. S.; Hager, C. S. Ethyl *n*-butylacetoacetate (Caproic and, α-acetyl-, ethyl ester). *Org. Synth. Collect*. **1941**, *1*, 248.
(107) Huckin, S. N.; Weiler, L. J. Alkylation of Dianions of *β*-keto Esters. *J. Am. Chem. Soc*. **1974**, *96*, 1082−1087.
(108) Ferraz, H. M. C.; Payret-Arrua, M. E.; De Oliveira, E. O.; Brandt, C. A. A New and Efficient Approach to Cyclic *β*-enamino-esters and *β*-enamino-ketones by Iodine-Promoted Cyclization. *J. Org. Chem*. **1995**, *60*, 7357−7359.
(109) Hermecz, I.; Keresztúri, G.; Vasvári-Debreczy, L. Aminomethylen-emalonates and Their use in Heterocyclic Synthesis. *Adv. Heterocycl. Chem*. **1992**, *54*, 1−429.
(110) Smrkovski, L. L.; Buck, R. L.; Alcántara, A. K.; Rodríguez, C. S.; Uylangco, C. V. Studies of Resistance to Chloroquine, Quinine, Amodiaquine and Mefloquine Among Philippine Strains of Plasmodium falciparum. *Trans. R. Soc. Trop. Med. Hyg*. **1985**, *7*, 37−41.

CI050085T