Available online at www.sciencedirect.com
SCIENCE DIRECT
European Journal of Medicinal Chemistry 41 (2006) 483–493
EUROPEAN JOURNAL OF MEDICINAL CHEMISTRY
http://france.elsevier.com/direct/ejmech

Original paper

Non-stochastic quadratic fingerprints and LDA-based QSAR models in hit and lead generation through virtual screening: theoretical and experimental assessment of a promising method for the discovery of new antimalarial compounds

Alina Montero-Torres [a,*], Rory N. García-Sánchez [b,e], Yovani Marrero-Ponce [a,c,d], Yanetsy Machado-Tugores [a], Juan J. Nogal-Ruiz [e], Antonio R. Martínez-Fernández [e], Vicente J. Arán [f], Carmen Ochoa [f], Alfredo Meneses-Marcel [a], Francisco Torrens [d]

[a] *Department of Drug Design, CBQ, Central University of Las Villas, 54830 Santa Clara, Villa Clara, Cuba*
[b] *Laboratorio de Investigación de Productos Naturales Antiparasitarios de la Amazonía, Universidad Nacional de la Amazonía Peruana, Pasaje Los Paujiles s/n, A A.H H Nuevo San Lorenzo, San Juan Bautista, Iquitos, Peru*
[c] *Department of Pharmacy, Faculty of Chemistry-Pharmacy, Central University of Las Villas, 54830 Santa Clara, Villa Clara, Cuba*
[d] *Institut Universitari de Ciència Molecular, Universitat de València, Edifici d'Instituts de Paterna, P.O. Box 22085, E-46071 Valencia, Spain*
[e] *Departamento de Parasitología, Facultad de Farmacia (UCM), Av. Complutense s/n E-28040 Madrid, Spain*
[f] *Instituto de Química Médica (CSIC), Juan de la Cierva 3, E-28006 Madrid, Spain*

Received in revised form 16 December 2005; accepted 20 December 2005
Available online 20 March 2006

Abstract

In order to explore the ability of non-stochastic quadratic indices to encode chemical information in antimalarials, four quantitative models for the discrimination of compounds having this property were generated and statistically compared. Accuracies of 90.2% and 83.3% for the training and test sets, respectively, were observed for the best of all the models, which included non-stochastic quadratic fingerprints weighted with Pauling electronegativities. With a comparative purpose and as a second validation experiment, an exercise of virtual screening of 65 already-reported antimalarials was carried out. Finally, 17 new compounds were classified as either active/inactive ones and experimentally evaluated for their potential antimalarial properties on the ferriprotoporphyrin (FP) IX biocrystallization inhibition test (FBIT). The theoretical predictions were in agreement with the experimental results. In the assayed test compound **C5** resulted more active than chloroquine. The current result illustrates the usefulness of the *TOMOCOMD-CARDD* strategy in rational antimalarial-drug design, at the time that it introduces a new family of organic compounds as starting point for the development of promising antimalarials.

Keywords: Antimalarial compounds; QSAR; LDA; Non-stochastic quadratic index; *TOMOCOMD-CARDD* software

1. Introduction

Being known since time immemorial, malaria remains a serious and complex health problem today. Approximately 300–400 million people worldwide suffer from this infectious disease, dying ca. 3 million every year, mostly children younger than 5 years [1–3]. This situation has become even more complex over the last decades with the increase in resistance to those drugs normally used to combat the malaria parasite [3–5]. Therefore, the international scientific community is called upon to use efficient strategies for the discovery of non-traditional antimalarial agents [6–9].

Computer-aided drug design has emerged in the pharmaceutical world as an important tool for the rational search of che-

* Corresponding author. Tel.: +53 42 28 1473; fax: +53 42 28 1130, +53 42 28 1455.
E-mail address: alinamontero@gmail.com (A. Montero-Torres).

0223-5234/$ - see front matter © 2006 Elsevier SAS. All rights reserved.
doi:10.1016/j.ejmech.2005.12.010

micals with desired properties. Different studies related to the in silico design have been reported in the literature during the last years [7–16]. In this context, our research group has recently introduced the *TOMOCOMD*-CARDD scheme for the codification of chemical information of small and medium-sized molecules considering different atomic features [17–23]. Using the *TOMOCOMD*-CARDD strategy, it is possible to encode "critical" structural patterns of data sets consisting of compounds with common molecular scaffolds as well as very diverse structures. Specifically, non-stochastic quadratic indices were defined upon the basis of linear algebra theory by one of the present authors, as the first descriptors family implemented in the *TOMOCOMD-CARDD* software [20]. These fingerprints have been successfully used in the generation of *quantitative structure–activity relationships* (QSAR) to predict physical and pharmacological properties of organic compounds [9,20,22,24–27]. In this context, biological properties like antibacterial [25], trichomonacidal [26] and trypanocidal activities [27] have been successfully studied. In an earlier publication, an overview of the significance of these *TOMOCOMD* descriptors and a comparison whit other molecular indices have been presented [22].

With the aim of extending the application field of the *TOMOCOMD*-CARDD method, the present study is intended to develop quantitative models to discriminate antimalarial compounds from inactive ones using, in a comparative way, four families of atomic labels. It is also an objective of the present report to conduct exercises of virtual screening for antimalarials of known activity as well as for compounds for which the activity of interest has been left undetected so far. Both experiments can be considered as part of the external validation process. Results of virtual evaluation and in vitro antimalarial activity of 17 already synthesized homocyclic and heterocyclic compounds will be presented.

2. Materials and methods

2.1. Computational approach

Calculations were carried out on a PC Pentium-4 2.0 GHz. The *CARDD* module implemented in the *TOMOCOMD* software [17] was used to calculate total and local non-stochastic quadratic indices for the data set of organic molecules reported by R. Gozalbes et al. [7]. Weighting schemes depicted in Table 1 were used as atomic labels (molecular vector's components) [18,19].

We calculated k^{th} non-stochastic total quadratic indices $\boldsymbol{q_k}(x)$ and $\boldsymbol{q_k}^{\text{H}}(x)$ (without and with consideration of H atoms, respectively) as quadratic forms as follows:

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

where n is the number of atoms in the molecule; $x_1,\ldots,x_n$ are the coordinates or components of the "molecular vector" (X) (which can be seen as weights for each kind of atom in the molecule), and ${}^{k}a_{ij}$ are the elements of the k^{th} power of the symmetric square matrix **M**(G) of the molecular pseudograph (G). This matrix is constructed following the spirit of an "extended Hückel" molecular orbital model and encloses information of all valence-bond electrons (σ- and π-networks). For powers greater than 1, the electronic interaction through the chemical network is also taken into account [20].

Table 1
Values of the atomic weights used for quadratic index calculation

ID	Atomic mass	VdW[a] volume	Polarizability	Pauling electronegativity
H	1.01	6.709	0.667	2.20
B	10.81	17.875	3.030	2.04
C	12.01	22.449	1.760	2.55
N	14.01	15.599	1.100	3.04
O	16.00	11.494	0.802	3.44
F	19.00	9.203	0.557	3.98
Al	26.98	36.511	6.800	1.61
Si	28.09	31.976	5.380	1.9
P	30.97	26.522	3.630	2.19
S	32.07	24.429	2.900	2.58
Cl	35.45	23.228	2.180	3.16
Fe	55.85	41.052	8.400	1.83
Co	58.93	35.041	7.500	1.88
Ni	58.69	17.157	6.800	1.91
Cu	63.55	11.494	6.100	1.90
Zn	65.39	38.351	7.100	1.65
Br	79.90	31.059	3.050	2.96
Sn	118.71	45.830	7.700	1.96
I	126.90	38.792	5.350	2.66

[a] VdW: van der Waals.

k^{th} local quadratic indices of heteroatoms (S,N,O) [$\boldsymbol{q_{kL}}(x_{\text{E}})$ and $\boldsymbol{q_{kL}}^{\text{H}}(x_{\text{E}})$], hydrogen bonded to heteroatoms (S,N,O) [$\boldsymbol{q_{kL}}(x_{\text{E-H}})$] and halogens [$\boldsymbol{q_{kL}}(x_{\text{H}})$], as fragments [20], were calculated as:

$$q_{kL}(x) = \sum_{i=1}^{n}\sum_{j=1}^{n} {}^{k}a_{ijL}x_ix_j$$

where ${}^{k}\boldsymbol{a}_{ijL}$ is the element of the "i" row and "j" column of $\mathbf{M}_{\mathbf{L}}^{\boldsymbol{k}}$. Matrix $\mathbf{M}^{\boldsymbol{k}}$ encodes information of the selected fragment as well as of the molecular environment.

2.2. Statistical analysis

Linear discriminant analysis (LDA) was performed with the STATISTICA 5.5 package for Windows [28]. Forward stepwise procedure was used for variable selection [28]. Quantitative models with the following form were obtained:

$$\begin{aligned}\mathrm{P} = {} & \mathrm{a_0q_0(x)} + \mathrm{a_1q_1(x)} + \ldots + \mathrm{a_nq_n(x)} + \mathrm{a_{n+1}q_0L(x)} \\ & + \mathrm{a_{n+2}q_1L(x)} + \ldots + \mathrm{a_mq_mL(x)}\end{aligned}$$

where $\boldsymbol{P}$ is the biological property, $\boldsymbol{q_n}(x)$, the n^{th} total quadratic index, $\boldsymbol{q_{mL}}(x)$, the m^{th} local quadratic index and a_n's and a_m's, the coefficients obtained by LDA. The antimalarial activity was coded by a dummy variable "***Class***", which indicates the presence of either an active compound (***Class*** = 1) or an inactive one (***Class*** = –1). The classification of all cases was per-

formed by means of the posterior classification probabilities. Each compound was classified as either active if $\Delta P\% > 0$ or inactive otherwise, being:

$$\Delta P\% = [P(\text{Active}) - P(\text{Inactive})] \times 100$$

P(Active) and P(Inactive) are the probabilities with which the equation classifies a compound as either active or inactive, respectively. Those compounds with $\Delta P\% < 5\%$ were considered as not classified (nc).

The quality of the models was primarily determined by examining the Wilks' λ parameter, which can take values in the range from 0 (perfect discrimination) to 1 (no discrimination), the square Mahalanobis distance D^2, which indicates the separation between the two groups, the Fisher ratio F, the corresponding p-level, as well as the ratio between cases and variables in the equations [28,29]. For each model, Matthews correlation coefficients, accuracy values, specificities, hit rates (sensitivities), and sensitivities for the negative category were computed for the training and test groups [30].

2.3. Virtual screening

An exercise of virtual screening was conducted as an extension of the model validation process. Two series containing 65 antimalarials [3,31] and 17 new homocyclic and heterocyclic compounds [32–38] were used as second and third external test groups, respectively. These compounds were never used in the generation of models. Non-stochastic quadratic fingerprints were computed for each molecule with the *CARDD* module of the *TOMOCOMD* software [17]. By means of the calculation of posterior probabilities using the STATISTICA package [28] and following the criteria explained in Section 2.2, each case was classified as either active if $\Delta P\% > 0$ or as inactive otherwise. Finally, the canonical analysis was performed for the best model as implemented in the STATISTICA software [28].

2.4. Chemistry

All the compounds were synthesized as reported by Arán and co-workers [32–38].

2.5. Ferriprotoporphyrin (FP) IX biocrystallization inhibition test (FBIT)

The procedure for testing FP biocrystallization was performed according to the method of Deharo et al. [39]. In a non-sterile flat bottom 96-well plate at 37 °C for 18–24 h, a mixture containing either 50 μl of drug solution (from 5 to 0.0125 mg/ml) or 50 μl of solvent (for control), 50 μl of 0.5 mg/ml of haemin chloride (Sigma H 5533) freshly dissolved in dimethyl sulfoxide (DMSO) and 100 μl of 0.5 M sodium acetate buffer pH 4.4 was placed. The final pH of the mixture was in the range 5.0–5.2. The following order of addition was followed: first the haemin chloride solution, second the buffer, and finally either the solvent or the drug solution. The plate was then centrifuged at $1600 \times g$ for 5 min. The supernatant was discarded by vigorous flipping of the plate upside down twice. The remaining pellet was resuspended with 200 μl of DMSO to remove unreacted FP. The plate was then centrifuged once again and the supernatant similarly discarded. The pellet, consisting of precipitate of β-hematin, was dissolved in 150 μl of 0.1 M NaOH for direct spectroscopic quantification at 405 nm with a micro-plate reader (ELX 800 Biotek Instruments, Inc.). The percentage of inhibition of FP biocrystallization was calculated as follows:

Inhibition (%) = 100 × [(O.D. control − O.D. drug)/O.D. control] where O.D. represents the mean of optical densities for either controls or drugs.

IC50 values were determined using the *TENDANCE* function of software Excel [40].

3. Results and discussion

3.1. Development of discriminant functions

LDA has been extensively used by different authors in rational design processes. This has been the selected statistical technique for the development of the already reported *TOMOCOMD-CARDD* approaches [25–27]. By combining variables, this pattern-recognition method allows distinguishing between two or more groups of populations. In the present report, this kind of study is carried out with a group consisting of 25 antimalarials and 34 compounds with other biological actions reported by Gozalbes et al. [7]. The results of two exploratory k-MCAs (k-means cluster analyses) [41,42] were in agreement with the distribution of cases in the training and test sets considered by R. Gozalbes et al. [7]. Therefore, the representativeness of the whole population is ensured in both series, allowing the evaluation of the performance for cases in the same data domain. 17 antimalarials and 24 "inactive" (non-antimalarial) compounds were selected for the training data set. The remaining structures were prepared as external test group.

Making use of the LDA technique implemented in the STATISTICA software [28], the four mathematical models shown in Table 2 were obtained. Table 3 summarizes the statistical parameters for each of them.

Table 2
Obtained models

$$\textbf{\textit{Class}} = -2.1024 - 0.1249 \times 10^{-7}\ {}^{AM}\boldsymbol{q}_{12L}(x_E) + 0.3291 \times 10^{-8}\ {}^{AM}\boldsymbol{q}_{13L}{}^{H}(x_E) - 1.2102 \times 10^{-3}\ {}^{AM}\boldsymbol{q}_{0L}(x_H) + 0.6116 \times 10^{-4}\ {}^{AM}q_{5L}(x_H) \quad (1)$$

$$\textbf{\textit{Class}} = -2.1746 - 1.1312 \times 10^{-6}\ {}^{P}\boldsymbol{q}_{12L}(x_E) + 0.2931 \times 10^{-6}\ {}^{P}\boldsymbol{q}_{13L}{}^{H}(x_E) - 0.5251\ {}^{P}\boldsymbol{q}_{0L}(x_H) + 1.0109 \times 10^{-2}\ {}^{P}q_{5L}(x_H) \quad (2)$$

$$\textbf{\textit{Class}} = -2.4383 - 0.7757 \times 10^{-8}\ {}^{V}\boldsymbol{q}_{12L}(x_E) + 0.2009 \times 10^{-8}\ {}^{V}\boldsymbol{q}_{13L}{}^{H}(x_E) - 0.6581 \times 10^{-2}\ {}^{V}\boldsymbol{q}_{0L}(x_H) + 0.9119 \times 10^{-4}\ {}^{V}\boldsymbol{q}_{5L}(x_H) \quad (3)$$

$$\textbf{\textit{Class}} = -3.7251 - 3.2721 \times 10^{-7}\ {}^{PE}\boldsymbol{q}_{12L}(x_E) + 8.6032 \times 10^{-8}\ {}^{PE}\boldsymbol{q}_{13L}{}^{H}(x_E) - 0.1246\ {}^{PE}\boldsymbol{q}_{0L}(x_H) + 3.7808 \times 10^{-3}\ {}^{PE}\boldsymbol{q}_{5L}(x_H) \quad (4)$$

Table 3
Overall measures of accuracy obtained for models **1**–**4**

Models [a]	Matthews correlation coefficient (*C*)	Accuracy (%)	Specificity (%)	Sensitivity 'hit rate' (%)	Sensitivity (–) (%)	Wilks' λ	D^2	F
Training Series								
1	0.74	85.4	92.3	75.0	95.8	0.52	3.64	8.3
2	0.75	85.4	92.9	76.5	95.6	0.55	3.23	7.4
3	0.75	87.8	92.9	76.5	95.8	0.54	3.40	7.8
4	*0.80*	*90.2*	*93.3*	*82.3*	*95.8*	*0.48*	*4.30*	*9.9*
Test Series								
1	0.55	77.8	83.3	62.5	90.0	0.52	3.64	8.3
2	0.47	66.7	66.7	66.7	80.0	0.55	3.23	7.4
3	0.53	72.2	75.0	75.0	77.8	0.54	3.40	7.8
4	*0.77*	*83.3*	*100.0*	*71.4*	*100.0*	*0.48*	*4.30*	*9.9*

[a] Models **1**–**4** are depicted in Table 2.

By using the obtained discriminant functions, each compound can then be classified according to the posterior probabilities as either active or inactive one ΔP% = [P(Active)–P(Inactive)] × 100. In Tables 4 and 5, the results using models **1**–**4** for the training and test series are presented.

Mathews correlation coefficient is a measure that provides a balanced evaluation of the prediction, because it uses the quantities of true positives, true negatives, false positives and false negatives [30]. This statistic coefficient was calculated for each set and each model. As it can be observed in Table 3, the model that shows the best MCC resulted the number **4**. We also consider as evaluating criterion the probability of correctly predicting a positive example (sensitivity or hit rate), the probability that a positive prediction will be correct (specificity) and the negative predictive value (sensitivity of the negative category), which gives a criterion of good classification for the inactive group. An appropriate discriminatory power for those compounds included in the training group was observed in all models. Active and inactive compounds were differentiated with an accuracy of 90.2% and 85.4% for the best and worst models, respectively (Table 3). However, models **1**, **2** and **3** presented very low predictive ability expressed in terms of accuracy and Mathews correlation coefficient (MCC) for the test group. On the contrary, model **4**, which was obtained using Pauling electronegativities as atomic weights, showed adequate values of accuracy, MCC, specificity and sensitivity for both categories

Table 4
Classification of active compounds in training and test data sets using models **1**–**4**

Chemicals	ΔP% [a]	Cl.	ΔP% [b]	Cl.	ΔP% [c]	Cl.	ΔP% [d]	Cl.
Training active group								
Acedapsone	99.54	+	99.56	+	99.70	+	99.79	+
Amodiaquine	96.43	+	95.77	+	96.44	+	98.30	+
β-Arteether	–49.71	–	–59.12	–	–62.50	–	–72.58	–
Artemisinin	–58.35	–	–57.55	–	–61.32	–	–77.73	–
Cinchonine	3.53	nc	5.38	+	13.79	+	6.17	+
Chloroguanide	70.07	+	86.42	+	82.76	+	71.51	+
Chloroquine	87.30	+	87.05	+	86.19	+	86.72	+
Chloroproguanil	98.84	+	99.29	+	99.00	+	98.55	+
Dapsone	96.40	+	95.96	+	97.49	+	97.01	+
Halofrantine	99.83	+	99.73	+	99.79	+	99.69	+
Methylarsacetin	–56.89	–	–38.55	–	–41.58	–	14.94	+
Mefloquine	83.87	+	71.29	+	88.17	+	92.54	+
Pamaquine	11.87	+	11.61	+	20.77	+	25.24	+
Pyrimetamine	72.01	+	71.48	+	61.89	+	70.12	+
Quinacrine	91.39	+	89.68	+	89.24	+	94.15	+
Quinine	27.18	+	18.86	+	29.78	+	42.15	+
Quinoline	–41.19	–	–36.73	–	–36.84	–	–63.54	–
Test active group								
β-Artemether	–50.14	–	–59.66	–	–63.05	–	–73.72	–
Artesunate	–45.97	–	–49.03	–	–51.60	–	–62.21	–
Bebeerine	70.41	+	53.15	+	64.95	+	94.41	+
Cycloguanil	–13.63	–	85.81	+	74.39	+	0.71	nc
Hydroxychloroquine	88.10	+	87.50	+	86.79	+	88.03	+
Plasmocid	10.95	+	10.21	+	19.31	+	21.72	+
Primaquine	5.90	+	4.19	nc	12.18	+	14.80	+
Quinocide	6.03	+	4.19	nc	12.22	+	14.64	+

[a,b,c,d] Results of the classification of active compounds in training and test sets obtained from models **1**–**4** using atomic masses, polarizabilities, van der Waals volumes, and Pauling electronegativities as atomic weights respectively. ΔP% = [P(Active) – P(Inactive)] × 100. nc = not classified.

Table 5
Classification of inactive compounds in training and test data sets using models **1**–**4**

Chemicals	ΔP% [a]	Cl.	ΔP% [b]	Cl.	ΔP% [c]	Cl.	ΔP% [d]	Cl.
Training inactive group								
Androsterone	–67.36	–	–69.93	–	–74.36	–	–88.08	–
Anisindione	–33.90	–	–44.62	–	–46.72	–	–41.19	–
Apronalide	–80.80	–	–63.87	–	–69.95	–	–92.82	–
Aspirin	–67.86	–	–64.40	–	–69.35	–	–82.83	–
Benomyl	–80.54	–	–3.99	nc	–11.14	–	–83.62	–
Benzphetamine	–64.33	–	–63.61	–	–67.41	–	–87.97	–
Bethanidine	–69.55	–	–58.68	–	–63.32	–	–91.24	–
Biguanide	–76.36	–	–52.51	–	–60.15	–	–87.71	–
Bixin	–60.21	–	–53.84	–	–56.79	–	–84.19	–
p,α-Dibromotoluene	–90.94	–	–85.78	–	–80.81	–	–52.44	–
Bucrylate	–91.74	–	–20.12	–	–30.02	–	–98.79	–
Butamisole	–5.50	–	34.04	+	34.52	+	–5.75	–
Butoxycaine	–44.27	–	–51.97	–	–54.66	–	–63.82	–
Capsaicin	–34.87	–	–41.24	–	–42.16	–	–53.23	–
Carbimazole	–90.08	–	–57.12	–	–58.53	–	–91.59	–
Carbuterol	–18.57	–	–21.05	–	–18.49	–	–25.47	–
Caroxazone	–50.01	–	–45.07	–	–47.37	–	–68.05	–
Cinoxacin	–68.67	–	–57.86	–	–63.92	–	–75.05	–
Cyromazine	–66.79	–	–63.77	–	–70.00	–	–76.69	–
Diethylpropion	–59.30	–	–62.54	–	–66.55	–	–80.73	–
Dimecrotic acid	–51.59	–	–53.32	–	–57.00	–	–63.88	–
Dithianone	–98.92	–	–96.99	–	–97.41	–	–99.48	–
Droxidopa	–24.81	–	–43.17	–	–43.46	–	–45.21	–
Fenoterol	12.02	+	–20.16	–	–15.14	–	8.45	+
Test inactive group								
Anthralin	–40.38	–	–53.77	–	–57.11	–	–52.69	–
Beclofen	46.56	+	43.45	+	59.39	+	–7.56	–
Barbital	–91.98	–	–47.66	–	–56.45	–	–97.19	–
β-Benzalbutyramide	–74.57	–	–71.34	–	–76.37	–	–90.84	–
Cefuroxime	–19.20	–	15.07	+	9.27	+	–9.33	–
Cotinine	–18.85	–	–9.08	–	–3.34	nc	–31.56	–
Cropropamide	–62.33	–	–51.94	–	–55.93	–	–79.10	–
Dopamine	–46.91	–	–58.59	–	–61.63	–	–72.85	–
Emodin	–23.07	–	–47.83	–	–50.46	–	–24.14	–
Emylcamate	–80.64	–	–74.87	–	–79.97	–	–94.46	–

[a,b,c,d] Results of the classification of inactive compounds in training and test sets obtained from models **1**–**4** using atomic masses, polarizabilities, van der Waals volumes, and Pauling electronegativities as atomic weights respectively. ΔP% = [P(Active) – P(Inactive)] × 100. nc = not classified.

(positives and negatives) on the evaluation of this group. Using model **4** on the training data group, just one inactive and three active compounds resulted missclassified (1/24, 3/17). In the test set, five compounds were correctly identified as actives and just one case could not be classified by the model. Furthermore, inactive observations were correctly classified.

As it can be observed in the training and test sets, model **4** showed very low false alarm rates (4.16% and 0.0%, respectively), meaning that we have just a few numbers of misclassified inactive compounds. This is a desirable condition for considering a model as adequate, if we take into account that this number represents inactive compounds that will be sent to biological assays and, in this way, loss of time and resources.

Considering that the quality of any QSAR depends on its ability to predict certain properties in compounds that were never used to obtain the classification functions, we conclude that model **4** should present more usefulness in the further identification of potential antimalarials. With the aim of illustrating the differences in predictive abilities between the four obtained models, antimalarial activity distribution diagrams were built [43]. Therefore, the discriminant functions were applied to the groups of active and inactive compounds selected for the present study. The expectancy E of finding a case in a given interval x is defined as either:

$$E_a = \frac{\text{Percentage of active compounds in } x}{\text{Percentage of inactive compounds in } x + 100}$$

$$E_i = \frac{\text{Percentage of inactive compounds in } x}{\text{Percentage of active compounds in } x + 100}$$

where E_a and E_i are the activity and inactivity expectancy respectively. In the intervals of minimum overlapping and greater E_a, the probability of finding new active compounds is the greatest. Fig. 1 shows the pharmacological distribution diagrams for each model (**1**–**4**). White and black bars represent active and inactive groups, respectively.

As it can be concluded from the analysis of these diagrams, the regions with minimum overlap for compounds with antimalarial properties are located for $0 < \text{Class} < 7$. As expected, model **4** showed the lowest overlap in the positive region.

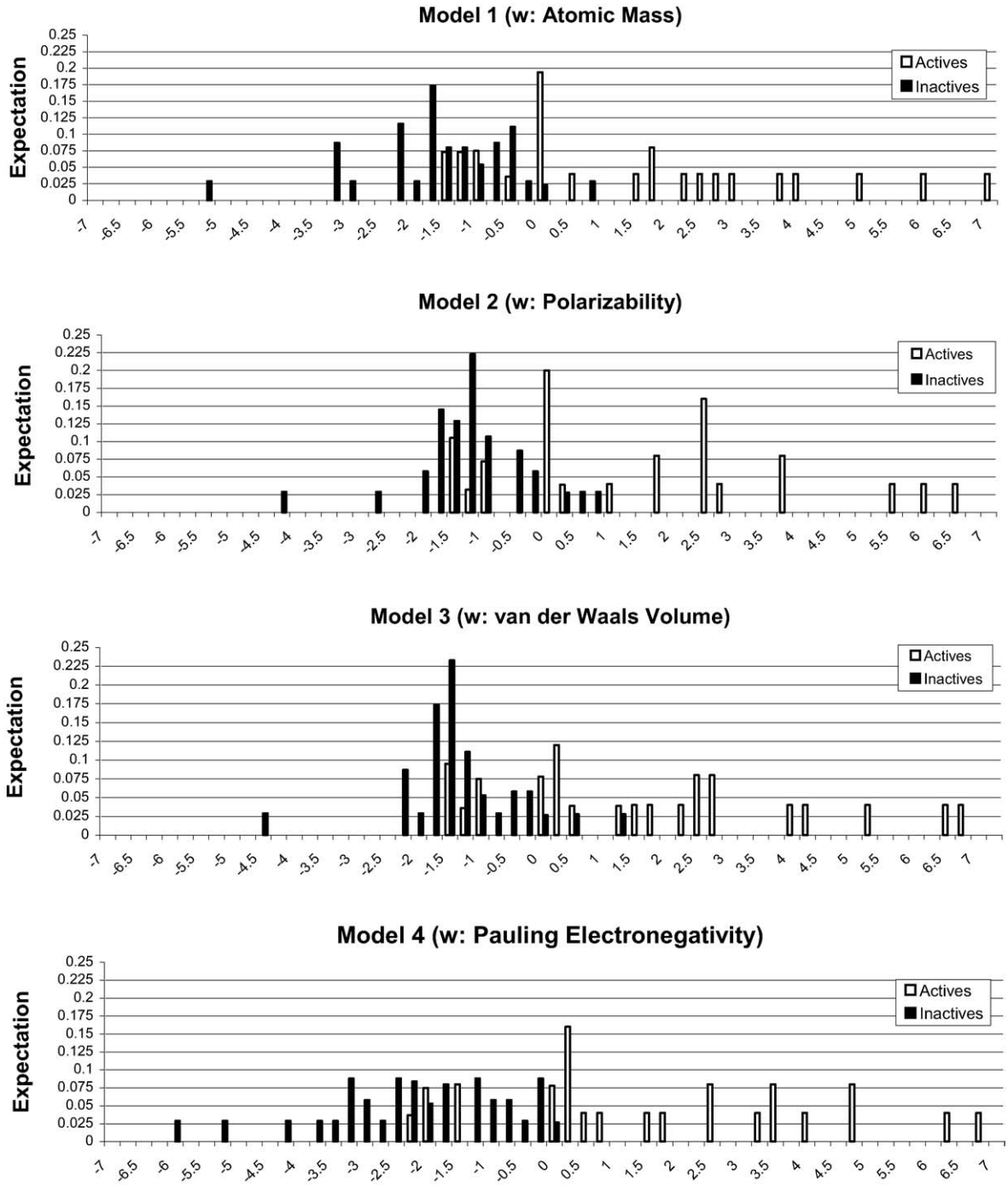

Fig. 1. Pharmacological distribution diagrams for antimalarial activity from the models **1**–**4**.

Making an analysis of the differences between the descriptors included in models **1**–**4**, we observe that the fingerprints chosen in the stepwise analysis for each model contain the same topological information. Nevertheless, they were computed considering different atomic labels and, in this sense, different kinds of atomic information. It is apparent that, in this case, critical aspects on the structures of active and inactive compounds are encoded more efficiently through electronegativity values than using atomic masses, polarizabilities or van der Waals volumes.

3.2. Identifying antimalarial compounds through virtual screening

In an attempt for extending the analysis of the predictive ability of all these models, a group of 65 antimalarials collected from the literature was evaluated [3,31]. In this case, this set of compounds can be considered as a second test group. The results of this preliminary virtual screening exercise are shown in Table 6. As it can be seen, model **4** yields, also here, the best

Table 6
Results of ligand-based virtual screening simulations

Chemicals	ΔP% [a]	Cl.	ΔP% [b]	Cl.	ΔP% [c]	Cl.	ΔP% [d]	Cl.	Ref.
Quinidine	27.18	+	18.86	+	29.78	+	42.15	+	[3]
Cinchonidine	3.53	nc	5.38	+	13.79	+	6.17	+	[3]
WR 122,455	84.24	+	66.46	+	85.82	+	82.25	+	[3]
WR 225498	74.53	+	62.40	+	78.10	+	90.07	+	[3]
WR 242511	0.19	nc	-4.09	–	1.61	nc	15.78	+	[3]
WR 238,605	68.65	+	51.54	+	68.82	+	87.68	+	[3]
CDRI 8053	14.46	+	39.80	+	49.63	+	52.83	+	[3]
WR182393	95.06	+	99.23	+	98.62	+	93.10	+	[3]
Trimethoprim	17.53	+	3.06	nc	10.42	+	43.99	+	[3]
Sulfadoxine	96.06	+	86.85	+	95.55	+	89.57	+	[3]
Sulfalene	97.73	+	92.65	+	97.84	+	95.40	+	[3]
WR 99210	99.54	+	99.90	+	99.75	+	99.53	+	[3]
Sulfisoxazole	94.66	+	86.33	+	95.40	+	69.67	+	[3]
Sulfamethoxazole	96.34	+	90.22	+	96.39	+	76.78	+	[3]
Clociguanil	91.60	+	98.84	+	97.36	+	90.10	+	[3]
PS-15	99.92	+	99.90	+	99.83	+	99.89	+	[3]
Nitroquine	9.81	+	99.83	+	99.82	+	99.90	+	[31]
Fluornemethanol	99.90	+	99.85	+	99.69	+	99.71	+	[31]
Tripiperakin	99.95	+	99.95	+	99.96	+	99.99	+	[31]
Bispyroquine	97.91	+	97.66	+	98.26	+	99.33	+	[31]
Gossypol	41.04	+	–9.49	-	-4.03	nc	69.97	+	[31]
Cycloquin	97.46	+	97.09	+	97.74	+	99.06	+	[31]
Benzonaphthyridine 7351	98.34	+	97.97	+	98.44	+	99.71	+	[31]
12278 R	99.97	+	99.98	+	99.98	+	100.0	+	[31]
Tebuquine	99.67	+	99.64	+	99.59	+	99.80	+	[31]
Octanoylprimaquine	7.51	+	13.71	+	22.37	+	25.27	+	[31]
Pyronaridine	9.64	+	98.54	+	98.97	+	99.77	+	[31]
Amopyroquine	96.74	+	96.14	+	96.82	+	98.52	+	[31]
Ciprofloxacin	95.01	+	87.37	+	96.14	+	99.90	+	[31]
Norfloxacin	99.87	+	99.39	+	99.91	+	100.0	+	[31]
Arteflene	83.52	+	66.88	+	86.90	+	89.04	+	[31]
Enpiroline	88.76	+	78.70	+	91.40	+	93.73	+	[31]
Fenozan - 50 F	82.62	+	37.69	+	74.48	+	98.31	+	[31]
Exifone	12.23	+	–30.24	–	–28.51	–	23.65	+	[31]
Piperaquine	99.91	+	99.91	+	99.93	+	99.98	+	[31]
Dioncophyline B	–3.06	nc	–23.21	–	–20.49	–	2.27	nc	[31]
Berberine	72.10	+	60.67	+	72.40	+	92.82	+	[31]
Licochalcone A	10.14	+	–21.16	–	–17.51	–	21.32	+	[31]
Hexalorxylol	40.82	+	–50.85	–	–98.29	–	62.99	+	[31]
Metachloridine	99.74	+	98.26	+	99.39	+	99.31	+	[31]
WR 135 403	96.67	+	97.43	+	92.91	+	91.00	+	[31]
Cilional	13.86	+	10.68	+	20.11	+	23.48	+	[31]
WR 10 488	99.13	+	99.26	+	99.25	+	99.47	+	[31]
WR 226 253	99.75	+	99.55	+	99.64	+	99.78	+	[31]
Oxychlorochin	88.96	+	–25.37	–	87.14	+	88.74	+	[31]
Antimalarine	10.95	+	87.70	+	19.31	+	21.72	+	[31]
CI–608	100.00	+	10.21	+	100.00	+	100.0	+	[31]
Brindoxime	100.00	+	99.66	+	99.09	+	99.53	+	[31]
Endochin	–25.78	–	–35.79	–	–36.91	–	–28.12	–	[31]
Fluoroquine	64.58	+	47.22	+	72.01	+	91.55	+	[31]
Pentaquine	7.77	+	7.27	+	15.65	+	18.97	+	[31]
Isopentachin	9.26	+	8.91	+	17.51	+	22.44	+	[31]
Dabekhin	0.87	nc	7.68	+	16.73	+	0.27	nc	[31]
Methylchloroquine	84.93	+	84.99	+	83.32	+	84.55	+	[31]
Floxacrine	98.98	+	97.56	+	98.36	+	99.77	+	[31]
RC- 12	95.82	+	79.81	+	69.93	+	80.28	+	[31]
Azamepacrine	89.76	+	88.83	+	87.77	+	94.32	+	[31]
Aecachinium	16.19	+	24.89	+	35.53	+	36.57	+	[31]
Mepacrine	84.74	+	82.42	+	79.75	+	85.13	+	[31]
Aristochin	90.49	+	90.94	+	95.53	+	99.12	+	[31]
Chalcone	99.29	+	99.24	+	99.09	+	99.34	+	[31]
Naphthol blue-black	97.59	+	44.17	+	98.62	+	96.79	+	[31]
Hydroxypiperaquine	99.92	+	99.92	+	99.93	+	99.99	+	[31]

(continued)

Table 6 *(continued)*

Chemicals	ΔP% [a]	Cl.	ΔP% [b]	Cl.	ΔP% [c]	Cl.	ΔP% [d]	Cl.	Ref.
Methotrexate	30.40	+	39.36	+	47.22	+	81.92	+	[31]
Aminopterin	33.86	+	41.42	+	49.84	+	82.93	+	[31]
% Accuracy	*92.3*		*86.2*		*89.2*		*95.4*		

[a,b,c,d] Results for the classification of compounds obtained from Eqs. **1–4** using atomic masses, polarizabilities, van der Waals volumes and Pauling electronegativities as scheme weight to calculate non-stochastic quadratic indices, respectively: ΔP% = [P(Active) – P(Inactive)] × 100. nc = not classified.

predictions. In this case, the antimalarial behavior of 95.4% of this group was correctly predicted.

The utility of these algorithms can just be confirmed if adequate pharmacological trials are performed to corroborate the expected activity for compounds of unknown activity. Following this idea, a ligand-based experiment of lead virtual generation was carried out. In ligand-based methods, similar compounds are assumed to produce similar effects. It means that, if one or more active chemicals are known, it is possible to search a database identifying molecules with similar properties. In this way, a family consisting of 17 already synthesized compounds was evaluated with the models [33–38]. Their structures are shown in Fig. 2.

Ferriprotoporphyrin IX biocrystallization inhibition test (FBIT) was used for the in vitro screening of this family in order to corroborate the theoretical predictions. During their digestion of host cell hemoglobin, intraerythrocytic malaria parasites produce large amounts of toxic ferriprotoporphyrin IX (FP). The inhibition of biomineralization of FP to β-hematin by antimalarial compounds underlies their action mode and, in this sense, it can be used to give a criterion of potential antimalarial character [39]. The global results for the selected family in both, in vitro and in silico screenings are depicted in Table 7.

From 17 compounds, seven cases (**C1**–**C7**) showed IC_{50} values lower than 5.00 mg/ml, resulting active in the biomineralization microassay. The remaining ten resulted inactive ones. These results were in concordance with the theoretical predictions of models **1**–**4**. In three cases it was obtained 76.5% of accuracy. In this assay, compound **C5** resulted more active than Chloroquine (see Table 7). The chemical core of this family, and specially the structure of compound **C5,** can be considered as an important starting point for the design of novel antimalarials. In this sense, new refining algorithms are needed for optimizing the pharmacological, toxicological and physicochemical properties, as well as to explain the mode of action of this new set of promising drugs.

When LDA analysis is applied to solve the two groups classification problem, two classification functions (for positive and negative observations) are always found. Usually, the global function obtained by considering the difference between these first two is reported in QSAR studies, in which the main objective is to make a differentiation between members of the two populations. However, using these functions, it is not possible to obtain bivariate activity maps because they are not orthogonal. In this case a dimensional reduction through canonical analysis is required [44,45]. In order to apply the LDA results to the prediction of potency on the new synthesized family, in the current study a canonical analysis was carried

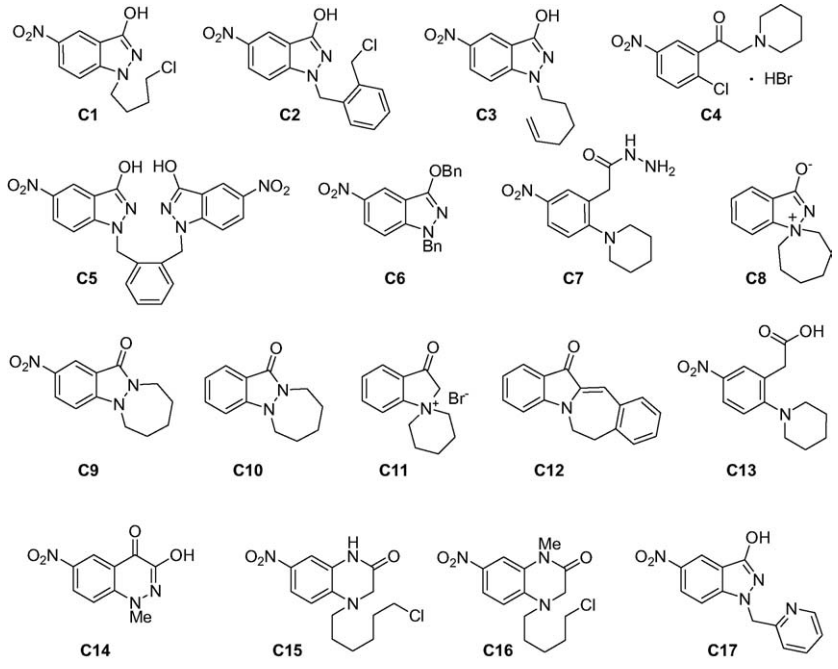

Fig. 2. Codes and structures of compounds screened in the current study.

Table 7
Screened compound, their in vitro antimalarial activity using chloroquine diphosphate as reference and their classification according to *TOMOCOMD-CARDD* approaches **1–4**

Compound	IC_{50} [mg/ml] ± S.D.	IC_{50} μM	Obs.	Predictions			
				1	*2*	*3*	*4*
Chloroquine	0.04	17.93	+	+	+	+	+
C1	2.06 ± 0.05	1910.05	+	–	–	–	–
C2	2.10 ± 0.34	1653.53	+	+	+	–	+
C3	1.00 ± 0.07	960.20	+	–	–	–	nc
C4	0.51 ± 0.10	351.91	+	+	+	+	+
C5	*0.03 ± 0.02*	*15.60*	+	+	+	+	+
C6	1.66 ± 0.13	1152.43	+	+	+	+	+
C7	1.65 ± 0.58	1485.21	+	–	nc	+	+
C8	> 5	> 5000	–	–	–	–	–
C9	> 5	> 5000	–	–	–	–	nc
C10	> 5	> 5000	–	–	–	–	–
C11	> 5	> 5000	–	–	–	–	–
C12	> 5	> 5000	–	–	–	–	–
C13	> 5	> 5000	–	–	–	–	–
C14	> 5	> 5000	–	–	–	–	–
C15	> 5	> 5000	–	–	–	–	–
C16	> 5	> 5000	–	–	–	–	–
C17	> 5	> 5000	–	+	+	+	+
% Accuracy				76.5	70.6	76.5	76.5

out by considering model **4**. In this sense a one-dimensional activity map was analyzed. In this experiment the canonical transformation yields the following canonical root:

$$\text{Root } 1 = -1.0981 - 1.5758 \times 10^{-7\ \mathrm{PE}} q_{12\mathrm{L}}(x_{\mathrm{E}}) + 0.414\,4 \times 10^{-7\ \mathrm{PE}} q_{12\mathrm{L}}{}^{\mathrm{H}}(x_{\mathrm{E}}) - 0.0608^{\mathrm{PE}} q_{0\mathrm{L}}(x_{\mathrm{H}}) + 1.8370 \times 10^{-3\ \mathrm{PE}} q_{5\mathrm{L}}(x_{\mathrm{H}}) \quad (5)$$

$N = 41$, $\lambda = 0.476$, $R_{can} = 0.723$, $\lambda^2 = 27.418$, P-level < 0.0001, Mean (+) = 1.2137, Mean(–) = –0.8597.

As a result of this kind of experiment, a sorting of compounds according to their activities can be presented. Analyzing the results depicted in Table 8, it should be noted that there is an overall rising trend of canonical scores, when they are plotted in the same order in which IC_{50} or log IC_{50} values decrease. In this sense a linear relationship between these groups of variables can be found. Fig. 3 illustrates the results of a linear regression analysis, which was conducted by considering canonical scores and log IC_{50} values.

In the previous case, the canonical analysis illustrates a second way in which the results of the present linear discriminant analysis can be used in the rational design of compounds with antimalarial potentialities.

Taking into account the results of the virtual screening and the biological behavior of the 17 assayed compounds, we can preliminarily conclude that an indazol core is preferred in order to achieve a positive biological reaction. However, this condition is not sufficient, because we observed compounds with an indazolic structure like **C17**, which does not present the desired property. Comparing specifically this case (**C17**) with other active indazols (**C1**, **C2**, **C3**, **C5** and **C6**) it can be observed that the attachment of a bulky (hydrophobic) moiety to Nitrogen atom number 1 (Fig. 2) can produce an enhancement in activity. Introduction of polar subfragments must be avoided. The presence of a Nitrogen atom in the aromatic ring attached to this position makes the compound inactive (see **C17**).

Table 8
Active synthesized compounds, their in vitro antimalarial activity and their classification according to the *TOMOCOMD-CARDD* approach **4**

Code	ΔP% [a]	Class [b]	Scores [c]	IC_{50} (μM)	log IC_{50} (μM)
C1	–42.40	–0.91	–0.27	1910.05	3.28
C2	48.05	1.05	0.68	1653.53	3.22
C3	–0.34	–0.01	0.17	960.20	2.98
C4	81.08	2.26	1.27	351.91	2.55
C5	98.97	5.26	2.71	15.60	1.19
C6	43.34	0.93	0.62	1152.43	3.06
C7	7.49	0.15	0.25	1485.21	3.17

[a] $\Delta P\% = [P(\text{Active}) - P(\text{Inactive})] \times 100$.
[b] Results for the classification of compounds obtained from Eq. (4).
[c] Canonical scores.

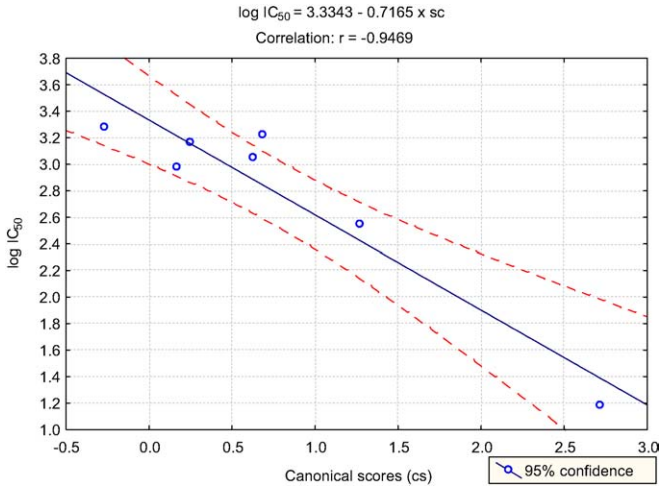

Fig. 3. Relationship between canonical scores and log IC_{50} (IC_{50} in μM) values for model **4**.

It was also observed that spiro and polycyclic molecules (like indols **C11** and **C12**, for example) resulted inactive ones. In structures having a halogen atom attached to a spacer longer than four atoms, it was also observed a lack of activity (see quinaxalines **C15** and **C16**). Compounds with a spacer consisting of four Carbons atoms (indazols **C1** and **C2**) resulted active ones but showed the greatest IC_{50} values. In order to analyze in a more exhaustive way the structure-activity relationships for this whole group, an extension of the studied set must be considered, including compounds with all possible substitution patterns. The dynamic and interactive character of the *TOMOCOMD* method allows updating the system by introduction of new observations and, in this sense, the generation of more and more refined approaches. Works on that direction are in progress.

4. Conclusions

The results of the present study show that, by an adequate choice of molecular descriptors and weighting schemes, it is possible to generate useful algorithms to identify the antimalarial character of new potential drugs. In this sense, *TOMOCOMD-CARDD* fingerprints have demonstrated their applicability in antimalarial recognition processes. The use of topological and electronic information, parametrized following the spirit of quadratic indices, and the introduction of electronegativities as complement allowed the generation of useful discriminant functions.

A group of seven new compounds with antimalarial potentialities were identified in this study. Further toxicological tests, as well as in vitro and in vivo assays, should be performed to permit the use of some of them (especially **C5**) or their derivatives in therapeutics. However, the identification of this family, making use of the *TOMOCOMD-CARDD* approach, constitutes an example of how this rational design technique can help reducing costs and increasing the rate in which new chemical entities progress through the pipeline.

Acknowledgements

Y.M.-P. thanks the program "Estades Temporals per a Investigadors Convidats" for a fellowship to work at Valencia University. F. T. acknowledges financial support from the Spanish MEC DGI (Project No. CTQ2004-07768-C02-01/BQU) and Generalitat Valenciana (DGEUI INF01-051 and INFRA03-047, and OCYT GRUPOS03-173).

References

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

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

[3] D.A. Castell, Burger´s Medicinal Chemistry and Drug Discovery.-5th ed. V.5, in: M.E. Wollf (Ed.), Wiley-Interscience, New York, 1997, pp. 4–91.

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

[5] A.J. Lin, A.B. Zikry, D.E. Kyle, 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. 40 (1997) 1396–1400.

[6] M.L. Go, T.L. Ngiam, A.L.C. Tan, K. Kuaha, P. Wilairat, Structure-Activity Relationships of Some Indolo(3,2-c)quinolines with Antimalarial Activity, Eur. J. Pharm. Sci. 6 (1998) 19–26.

[7] R. Gozalbes, J. Gálvez, A. Moreno, R. García-Domenech, Discovery of New Antimalarial Compounds by Use of Molecular Connectivity Techniques, J. Pharm. Pharmacol. 52 (1999) 111–117.

[8] Y. Marrero-Ponce, A. Montero-Torres, C. Romero-Zaldivar, I. Iyarreta-Veitía, M. Mayón Peréz, R. García Sánchez, Non-Stochastic and Stochastic Linear Indices of the "Molecular Pseudograph's Atom Adjacency Matrix": Application to "*in silico"* Studies for the Rational Discovery of New Antimalarial Compounds, Bioorg. Med. Chem. 13 (2005) 1293–1304.

[9] Y. Marrero-Ponce, M. Iyarreta, A. Montero-Torres, C. Romero, C.A. Brandt, P.E. Ávila, K.A. Kirchgatter, Y. Machado, Ligand-Based Virtual Screening and *in silico* Design of New Antimalarial Compounds Using Non-Stochastic and Stochastic Total and Atom-type Quadratic Maps, J. Chem. Inf. Model 45 (2005) 1082–1100.

[10] L.B. Kier, L.H. Hall, The Nature of Structure-Activity relationships and Their Relation to Molecular Connectivity, Eur. J. Med. Chem. 12 (1977) 307–312.

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

[12] E. Estrada, A. Peña, R. García-Domenech, Designing Sedative/Hypnotic Compounds from a Novel Substructural Graph-theoretical Approach, J. Comput. Aided Mol. Des. 12 (1998) 583–595.

[13] E. Estrada, A. Peña, In Silico Studies for the Rational Discovery of Anticonvulsant Compounds, Bioorg. Med. Chem. 8 (2000) 2755–2770.

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

[15] H. González-Díaz, E. Uriarte, Biopolimer stochastic Moments. I. Modeling Human Rhinovirus Cellular Recognition with Protein Surface Electrostatic Moments, Biopolymers 77 (2005) 296–303.

[16] H. González-Díaz, E. Tenorio, N. Castañedo, L. Santana, E. Uriarte, 3D QSAR Markov Model for Drug-induced Eosinophilia-Theretical Prediction and Preliminary Experimental Assay of the Antimicrobial Drug G1, Bioorg. Med. Chem. 13 (2005) 1523–1530.

[17] Y. Marrero-Ponce, V. Romero, (2002) TOMOCOMD software. Central University of Las Villas. TOMOCOMD (TOpological MOlecular COMputer Design) for Windows, version 2.0. This version can be obtained upon request to Y. Marrero: yovanimp@qf.uclv.edu.cu or ymarrero77@yahoo.es .

[18] R. Todeschini, P. Gramatica, New 3D Molecular Descriptors: The WHIM theory and QSAR Applications, Perspect. Drug Discov. Des. 9-11 (1998) 355–380.

[19] L. Pauling, in: The Nature of the Chemical Bond, Cornell University Press, Ithaca, N.Y, **1939**, pp. 2–60.

[20] Y. Marrero-Ponce, Total and Local Quadratic Indices of the Molecular Pseudograph's Atom Adjacency Matrix: Applications to the Prediction of Physical Properties of Organic Compounds, Mol. 8 (2003) 687–726.

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

[22] Y. Marrero-Ponce, Total and Local (Atom and Atom-Type) Molecular Quadratic Indices: Significance-Interpretation, Comparison to Other Mo-

lecular Descriptors and QSPR/QSAR Applications, Bioorg. Med. Chem. 12 (2004) 6351–6369.

[23] Y. Marrero-Ponce, H. González-Díaz, V. Romero-Zaldivar, F. Torrens, E.A. Castro, 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. 12 (2004) 5331–5342.

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

[25] Y. Marrero-Ponce, R. Median-Marrero, F. Torrens, Y. Martinez, V. Romero-Zaldivar, E.A. Castro, atom, Atom-type, and Total Non-Stochastic and Stochastic Quadratic Fingerprints: A Promising Approach for Modeling of Antibacterial Activity, Bioorg. Med. Chem. 13 (2005) 2881–2899.

[26] A. Meneses-Marcel, Y. Marrero-Ponce, Y. Machado Tugores, A. Montero-Torres, D. Montero-Pereira, J.A. Escario, J.J. Nogal-Ruiz, C. Ochoa, V.J. Arán, A.R. Martínez-Fernández, R.N. García-Sánchez, A Linear Discrimination Analysis Based Virtual Screening of Trichomonacidal Lead-Like Compounds. Outcomes of in silico Studies Supported by Experimental Results, Bioorg. Med. Chem. Lett. 17 (2005) 3838–3843.

[27] A. Montero-Torres, M.C. Vega, Y. Marrero-Ponce, M. Rolon, A. Gomez-Barrio, J.A. Escario, V.J. Aran, A.R. Martinez-Fernandez, A. Meneses-Marcel, A Novel Non-stochastic Quadratic Fingerprints-based Approach for the *in silico* Discovery of New Antitrypanosomal Compounds, Bioorg. Med. Chem. 13 (2005) 6264–6275.

[28] STATISTICA, version 5.5; StatSoft Inc., 1999.

[29] E. Estrada, In Topological Indices and Related Descriptors in QSAR and QSPR, in: J. Devillers, A.T. Balaban (Eds.), Gordon and Breach, Amsterdam, **1999**, pp. 403–453.

[30] P. Baldi, S. Brunak, Y. Chauvin, C.A. Andersen, H. Nielsen, Assessing the Accuracy of Prediction Algorithms for Classification: an Overview, Bioinformatics 16 (2000) 412.

[31] M. Negwer, Organic-Chemical Drugs and their Synonyms, Akademie-Verlag, Berlin, 1987.

[32] V.J. Arán, J.L. Asensio, J.R. Ruiz, M. Stud, The Heterocyclization of *N*′, *N*′-Disubstituted 2-Halogenobenzohydrazides to 1,1-Disubstituted Indazol-3-ylio Oxides, J. Chem. Res. (M) (1993) 1322–1345.

[33] V.J. Arán, J.L. Asensio, J.R. Ruiz, M. Stud, Reactivity of 1,1-Disubstituted Indazol-3-ylio Oxides: Synthesis of Some Substituted Indazolols and Indazolinones, J. Chem. Soc., Perkin Trans. 1 (1993) 1119–1127.

[34] J.R. Ruiz, V.J. Arán, J.L. Asensio, M. Flores, M. Stud, Synthesis of Quaternary Indoxyl Derivatives by Intramolecular Cyclization of Some Substitute Acetophenones, Liebigs Ann. Chem. (1994) 679–684.

[35] V.J. Arán, M. Flores, P. Muñoz, J.R. Ruiz, P. Sánchez-Verdú, M. Stud, Cytostatic Activity Agains HeLa Cells of a Series of Indazole and Indole Derivates; Synthesis and Evaluation of Some Analogues, Liebigs Ann. Chem. (1995) 817–824.

[36] V.J. Arán, M. Flores, P. Muñoz, J.A. Páez, P. Sánchez-Verdú, M. Stud, Analogues of Cytostatic, Fused Indazolinones: Synthesis, Conformational Análisis and Cytostatic Activity Agains HeLa Cells of Some 1-Substituted Indazolols, 2-Substitutes Indazolinones, and Related Compounds, Liebigs Ann. Chem. (1996) 683–691.

[37] V.J. Arán, J.L. Asensio, J. Molina, P. Muñoz, J.R. Ruiz, M. Stud, Approaches to 1,1-Disubstituted Cinnolin-3-ylio Oxides: Synthesis and Reactivity of a New Class of Heterocyclic Betaines, J. Chem. Soc., Perkin Trans. 1 (1997) 2229–2235.

[38] S. de Castro, R. Chicharro, V.J. Arán, Synthesis of Quinoxaline Derivatives from Substituted Acetanilides Through Intramolecular Quaternization Reactions, J. Chem. Soc., Perkin Trans. 1 (2002) 790–802.

[39] E. Deharo, R. Garcia, P. Oporto, M. Sauvain, P. Gautret, H.A. Ginsburg, Non-radiolabeled Ferriprotoporphyrin IX Biomineralization Inhibition Test (FBIT) for the High Throughput Screening of Antimalarial Compounds, Exp. Parasitol. 100 (2002) 252–256.

[40] Microsoft Office 2000 Professional, MicroSoft, 2000.

[41] J.W. Mc Farland, D.J. Gans, In Chemometric Methods in Molecular Design, in: H. van de Waterbeemd (Ed.), VCH Publishers, Weinheim, 1995, pp. 295–307.

[42] R.A. Johnson, D.W. Wichern, Applied Multivariate Statistical Analysis: Prentice-Hall, Englewood Cliffs, N.J, 1988.

[43] C. Calabuig, G.M. Antón-Fos, J. Gálvez, R. García-Doménech, New Hypoglycaemic Agents Selected by Molecular Topology, Int. J. Pharm. 278 (2004) 111–118.

[44] H. Van de Waterbeemd, In Chemometric Methods in Molecular Design, in: H. van de Waterbeemd (Ed.), VCH Publishers, Weinheim, **1995**, pp. 265–282.

[45] M.G. Ford, D.W. Salt, In Chemometric Methods in Molecular Design, in: H. van de Waterbeemd (Ed.), VCH Publishers, Weinheim, **1995**, pp. 283–292.