Pattern recognition methods investigation of ellipticines structure–activity relationships

Louraine C. de Melo [a,b], Scheila F. Braga [c], P.M.V.B. Barone [b,*]

[a] Centro Brasileiro de Pesquisas Físicas, Rua Dr. Xavier Sigaud, 150, 22290-180 Rio de Janeiro, RJ, Brazil
[b] Departamento de Física, Instituto de Ciências Exatas, Universidade Federal de Juiz de Fora, 36036-330 Juiz de Fora, MG, Brazil
[c] Instituto de Física "Gleb Wathagin", Universidade Estadual de Campinas, 13081-970 Campinas, SP, Brazil

Abstract

Ellipticine is a molecule derived from the natural extract *Ochrosia elliptica*. This molecule and its derivatives are highly cytotoxic to malignant cultured cells. The relatively simple structure of ellipticine has prompted chemists to design various structural modifications in order to obtain either more active derivatives or information on the structural moieties required for pharmacological activities. In the present work we report theoretical structure–activity relationship studies for 40 ellipticine derivatives using pattern-recognition methods such as electronics indices methodology (EIM), principal component analysis (PCA) and hierarchical clustering analysis (HCA) with molecular descriptors obtained from semiempirical parametric method 3 (PM3) calculations. By applying selected molecular descriptors it was possible to classify active and inactive compounds with accuracy up to 92% and also to suggest the activity of new untested molecules. These descriptors have been only recently discussed in the literature as new possible universal parameters for defining the biological activity of several classes of compounds.
© 2006 Elsevier Inc. All rights reserved.

Keywords: Ellipticine; Electronic indices methodology; Hierarchical clustering analysis; Olivacine; Principal component analysis; Semiempirical methods; Structure–activity relationships

1. Introduction

The structure of the plant alkaloid ellipticine (5,11-dimethyl-6*H*-pyrido[4,3]carbazole) which was reported in 1959 [1] is composed of a carbazole moiety linked to a pyridinic ring (Fig. 1). Anticancer studies of natural products led to the discovery of the antitumor properties of this compound and some of its derivatives [2].

Cancer chemotherapy consists in the administration of drugs which are supposed to kill malignant cells selectively in patients. Selective toxicity of the drugs toward malignant cells is required for significant antitumor activity resulting from cancer chemotherapy in either experimental models or in humans. This selective effect may come from either the occurrence of a specific damage to malignant cells or from relative less damage in normal cells compared to that in

malignant cells. The modifications of the primary target or changes in the accessibility of primary or ultimate targets are features that may account for the modulation of the cytotoxic efficacy of drugs in different cell lines [3].

In situations where DNA is the primary target for antitumor agents, the nature of the interaction between the drugs and DNA is likely to be the parameter that may control their cytotoxic efficacy. DNA binding by drugs such as ellipticines may occur through intercalation between base pairs by means of hydrophobic interactions, hydrogen bonding and van der Vaals forces [4]. The same physical processes are responsible for the outside binding of these drugs to either the minor or the major groove. External binding to DNA by means of electrostatic interactions has also been considered [4]. The hypothesis that intercalation of ellipticines between DNA base pairs could be responsible for antitumor activity has been supported by the fact that in this series, all drugs unable to intercalate exhibited a low cytotoxicity on cultured cells and no significant antitumor activity [5]. Intercalation binding of ellipticines in DNA has been observed by means of biochemical [6], thermodynamical

* Corresponding author. Tel.: +55 32 3229 3307; fax: +55 32 3229 3312.
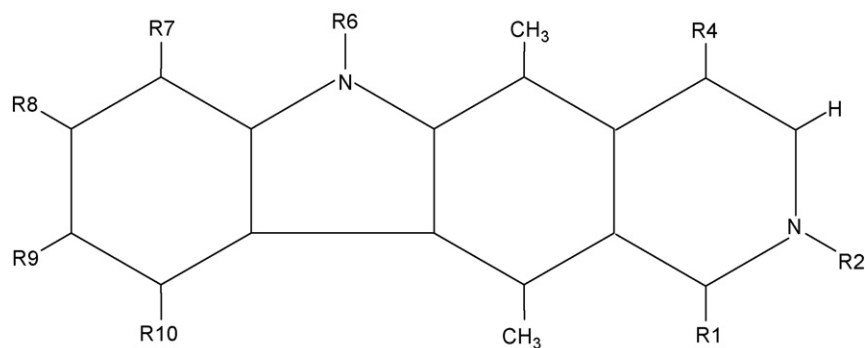 *E-mail address:* barone@fisica.ufjf.br (P.M.V.B. Barone).

Fig. 1. Basic structure of ellipticine.

[7,8] and spectroscopic techniques [6,9,10]. Recently Stiborová and coworkers have shown that enzymatic activated ellipticines can covalently bind DNA [11,12].

The search for the critical parameters involved in a given biological effect is the most commonly used approach in the experimental structure–activity relationship studies. This approach requires (i) the synthesis of many derivatives in a homologous series in which, ideally, only one parameter is varying and (ii) the use of appropriate biological models. Ellipticine derivatives represent useful samples for such a purpose because lots of synthetic compounds of this group have been prepared and their structure–activity relationships have been screened [13–16]. In the last 10 years a great number of new synthetic routes have been proposed to ellipticines which encourage further investigation of active derivatives [15–20]. The main reasons for interest in ellipticines for clinical purposes are their limited toxic side effects and their complete lack of hematological toxicity [21].

The possibility of having a large number of compounds opens a good perspective for the development of more efficient drugs. Structure–activity studies and biochemical testing of a large number of compounds are time and money consuming and might determine the commercial viability of a drug. Thus, if we could theoretically indicate *a priori* the most promising molecules as well as the inactive ones, this would be very useful in the design and development of new and better drugs [22–27].

Theoretical investigation of the molecular properties of ellipticine derivatives by means of electronic structure calculations reported in a previous work showed a direct correlation between the dipole moment values and molecular electrostatic potential (MEP) with the antitumor activity [28–30]. There are important differences in charge distribution among the molecules due to the presence of specific side groups, giving rise to remarkable differences in their dipole moments and MEP pattern (distribution of actives sites through the molecular skeleton) which provide criteria to classify molecules according to their antitumor activity.

The results of these previous studies [28–30] show that a global molecular parameter (the molecular dipole moment) and a property associated with a specific molecular region (the presence of active sites in the pyridinic ring) are important to define antitumor activity of these molecules. Thus it is worth to further investigate structure–activity relationships for ellipticines

by means of other global and local molecular electronic indices [31]. In this way the experimental evidence of the importance of substitutions at certain molecular sites [5] can be taken into account.

In this work we investigate the structure–activity relationships using the physicochemical descriptors as well as the theoretical approach based on the concepts of local density of electronic states (LDOS) developed in the electronic indices methodology (EIM) [31,32]. We also applied the multivariate pattern-recognition method principal component analysis (PCA) to select the best set of parameters that correlates with the ellipticine experimental antitumor activity. The overall similarity among molecules is calculated with the hierarchical cluster analysis (HCA) on the basis of the PCA pattern.

Our results show that observing theoretical descriptors, especially the EIM electronic ones, the activity of ellipticines can be monitored. EIM, PCA and HCA reproduced experimental classification with accuracy higher than 80%. The rules and patterns obtained in this investigation are applied to propose the activity of untested ellipticine derivatives.

2. Methodology

The basic structure of 40 ellipticine derivatives and the analogue compounds olivacine and isoellipticines included in this work is shown in Fig. 1 and Table 1. Olivacines, E32 to E37, differ from ellipticine only in the change of a methyl group from position R11 to position R1. Isoellipticines, E38 to E40, are linear isomeric ellipticines with a nitrogen atom in a different position in the pyridinic ring—see Table 1. In the present study we investigated the 40 molecules in their neutral forms.

For our studies we divided the set of 40 molecules into two groups: G1 and G2. The G1 group consists of the 25 ellipticine derivatives for which the biological activity is known (molecules 1–20 and 32–36 in Table 1). The G2 group consists of 15 ellipticine derivatives for which no experimental data are available and is considered for prediction purposes (molecules 21–31 and 37–40 in Table 1). Since we do not have experimental data under the same conditions for all the compounds studied, we applied in the G1 group the approach proposed by Villemin et al. [33] This approach consists of simply classifying the compounds into two classes: active (A) and inactive (I). To define active and

Table 1
Ellipticine and the derivatives studied

No.	Molecule	R_1	R_2	R_4	R_6	R_7	R_8	R_9	R_{10}	Act.
E01	Ellipticine (**E**)	H	–	H	H	H	H	H	H	A
E02	9E-Amino**E**	H	–	H	H	H	H	NH_2	H	I
E03	9-Hydroxi**E**	H	–	H	H	H	H	OH	H	A
E04	7-Hydroxi**E**	H	–	H	H	OH	H	H	H	I
E05	9-Metoxi**E**	H	–	H	H	H	H	OCH_3	H	A
E06[a]	Ellipticinium (**EM**)	H	CH_3	H	H	H	H	H	H	A
E07[a]	2N-Metil, 9-hydroxi**E**	H	CH_3	H	H	H	H	OH	H	A
E08[a]	2N-Metil, 1-metil, **E**	CH_3	CH_3	H	H	H	H	H	H	A
E09[a]	2N-Metil, 6N-metil, **E**	H	CH_3	H	CH_3	H	H	H	H	A
E10[a]	2N-Metil, 6N-metil, 9-hydroxi**E**	H	CH_3	H	CH_3	H	H	OH	H	A
E11	9-Hydroxi-6-metil**E**	H	–	H	CH_3	H	H	OH	H	A
E12	6-Metil**E**	H	–	H	CH_3	H	H	H	H	A
E13	9-Metoxy-1-amino**E**	NH_2	–	H	H	H	H	OCH_3	H	A
E14	9-Bromo**E**	H	–	H	H	H	H	Br	H	I
E15[a]	7-Hydroxy-2-metil**EM**	H	CH_3	H	H	OH	H	H	H	I
E16[a]	9-Metoxy-2-metil**EM**	H	CH_3	H	H	H	H	OCH_3	H	A
E17[a]	9-Hydroxy 7-metil**EM**	H	CH_3	H	H	CH_3	H	OH	H	A
E18[a]	9-Hydroxy 8,10-dimetil**EM**	H	CH_3	H	H	H	CH_3	OH	CH_3	I
E19[a]	9-Hydroxy-2-(2detilamino)etil**EM**	H	DEAEt	H	H	H	H	OH	H	A
E20[a]	2-Etil-9-hydroxy**E**	H	CH_2CH_3	H	H	H	H	OH	H	A
E21	9-Metoxy-6-metil**E**	H	–	H	CH_3	H	H	OCH_3	H	–
E22	11-Demetil-1-CHO-9-metoxy**E**	CHO	–	H	H	H	H	OCH_3	H	–
E23	11-Demetil-1-CH_2OH-9-metoxy**E**	CH_2OH	–	H	H	H	H	OCH_3	H	–
E24	6-Etil**E**	H	–	H	CH_2CH_3	H	H	H	H	–
E25	9-Metil**E**	H	–	H	H	H	H	CH_3	H	–
E26	7-Metil**E**	H	–	H	H	CH_3	H	H	H	–
E27	11-Demetil-9-hydroxy**E**	H	–	H	H	H	H	OH	H	–
E28[a]	4-HO-9-metoxy-2-metil**EM**	H	CH_3	HO	H	H	H	OCH_3	H	–
E29[a]	4-CO_2CH_3-9-metoxy-2-metil**EM**	H	CH_3	CO_2CH_3	H	H	H	OCH_3	H	–
E30[a]	9-Metil-2-metil**EM**	H	CH_3	H	H	H	H	CH_3	H	–
E31[a]	2-CH_2OCH_3**E**	H	CH_2OCH_3	H	H	H	H	H	H	–
E32	Olivacine (**O**)	CH_3	–	H	H	H	H	H	H	I
E33	7-Hydroxy**O**	CH_3	–	H	H	OH	H	H	H	I
E34	9-Hydroxy**O**	CH_3	–	H	H	H	H	OH	H	A
E35[a]	9-Hydroxy-2-metil**OM**	CH_3	CH_3	H	H	H	H	OH	H	A
E36[a]	9-Hydroxy-2,8,10-trimetil**OM**	CH_3	CH_3	H	H	H	CH_3	OH	CH_3	A
E37	9-Metoxy**O**	CH_3	H	H	H	H	H	OCH_3	H	–
E38	3N-Iso**E**[b]	H	H	H	H	H	H	H	H	–
E39	1N-11Demetil-2,4-dimetil**E**[b]	H	CH_3	CH_3	H	H	H	H	H	–
E40	4N-**E**[b]	H	H	–	H	H	H	H	H	–

The radical numbering refers to the structure shown in Fig. 1.

[a] Electronic open shell molecules.

[b] The number preceding the symbol N indicates the position of nitrogen atom in the pyridinic ring. A and I refer to active and inactive, **E** and **EM** refer to ellipticine and ellipticinium, and **O** and **OM** refer to olivacine and olivacinium.

inactive compounds for molecules E01 to E10 we consider the experimental scale of activity proposed by Cavaliere et al. [34]. For the remaining molecules the experimental index IC_{50} [2] has been considered. A molecule is considered active if the IC_{50} is about or lower than 1 μM. The IC_{50} index is defined as the drug concentration required for inhibition of 50% of the cancer cell reproduction rate, thus high IC_{50} values means low antitumor activity and vice versa.

We started our theoretical analysis by carrying out a fully geometrical optimization using the Mopac2000 program in CAChe package [35]. In order to find the energy global minimun we considered a detailed dihedral variation for the substituted chains in the ellipticine derivatives. All the calculations have been done with the PM3 semiempirical method. The selection of PM3 method for this investigation is based in our previous studies [30]. The PM3 Hamiltonian is a Hartree–Fock method based on the linear combination of atomic orbitals (LCAO) approximation. PM3 treats the one-center, two-electron integrals as pure parameters [36]. Unrestricted Hartree–Fock method was used for all calculations concerning open shell species.

Once the optimized geometries, eigenvalues, and eigenvectors had been obtained we proceeded into the calculations of electronic parameters of EIM. These parameters have been employed with success in classificatory problems of different molecular systems, including carcinogens [31,37,38], antitumor and antibiotics [39,40], hormones [41,42] and protease inhibitors compounds [43].

The electronic density of states (DOS) [44] is defined as the number of electronic states per energy unit. The related concept

of LDOS, i.e. the DOS calculated over a specific molecular region, is introduced in order to describe the spatial distribution of these states as well. These concepts can give us detailed information on the contributions of specific geometrical regions of the molecules to the chemical reactivity, optical response, etc., and, consequently, to their biochemical behavior.

For the LDOS calculations, the contribution of each atom to an electronic level is weighted by the square of the (real) molecular orbital coefficient, i.e. by the probability density corresponding to the level in that site. The summation is carried over the selected molecular regions. In other words, LDOS shows how an atom or a set of them contributes to the specified molecular orbitals [32].

One major advantage of the LDOS calculations is to allow the simultaneous determination of the relative contribution to each molecular level of the chosen molecular regions or the contribution of all separated regions to a specific molecular orbital (specially the frontier orbitals).

The energy separation of two molecular levels (Δ) and the difference between contributions of a molecular region to these electronic levels, obtained by LDOS (η), are the basis of the electronic indices methodology (EIM). Δ contains global molecular information, while η contains local molecular information and its is believed to reflect (on average) the relevant biochemical mechanisms [32].

In the present work we have collected and analyzed the LDOS contribution for many molecular regions of ellipticine skeleton including the ring presenting the highest bond order (region rA); the atomic sites N2 and N6 (regions rB and rC, respectively); the C_9 position (region rD); the pentagonal ring of the skeleton (region rF), the ring containing the C9 carbon (region rN), and the atomic sites 1–4 belonging to the pyridinic ring (region rP) (see Fig. 2). Each region has been chosen according to a particular motivation. Region rA is the most susceptible to electrophilic attack. For regions rB and rC the charge density associated with a DNA affinity (K_{app}) is correlated to antitumor and cytotoxic activities [28]. rD represents the site where substitutions are very relevant to the activity of ellipticines [5]. Region rF is the pentagonal ring of the structure where the heteroatom N is preserved in all the derivatives. Region rN represents one of the exposed rings of the skeleton (the other one being the region rA) where

biochemical reactions can occur, and rP corresponds to the region selected by MEP investigation as indicative of biologically inactive ellipticines [30].

The following calculated physicochemical descriptors have been chosen to be used in our structure–activity relationship analysis: the HOMO (highest occupied molecular orbital) and HOMO-1 energy; the energy difference between HOMO and HOMO-1 (ΔH); the LUMO (lowest unoccupied molecular orbital energy) and LUMO + 1 energy; the energy difference between LUMO and LUMO + 1 (ΔL); the contributions of specific molecular regions to the local density of states (LDOS) of HOMO and HOMO-1 levels (CH and CH-1) and their differences (ηH); the contributions of specific molecular regions to the local density of states (LDOS) of LUMO and LUMO + 1 levels (CL and CL + 1) and their differences (ηL); the dipole moment values (DM) and heat of formation (HF); refractivity (R), polarizability (Po), mass (M), volume (V), hydration (HE) and solvatation energies (SE) and coefficient of molecular partition octanol–water ($\log P$).

The above quantum chemical descriptors were obtained directly from the PM3 calculations and the $\log P$ was calculated by using parameters of the substituents' hydrophobicity [45].

The first eighth parameters listed above have been introduced in structure–activity relationship analysis by EIM. This methodology has been applied to identify active and inactive molecules from families of structurally related compounds [31,32,39–43,46]. EIM has been compared to more standard structure–activity relationship (SAR) multi-variate statistical analysis methods such as principal component analysis (PCA) and hierarchical clustering analysis (HCA) [47,48]. The results of EIM analysis and the comparison to other methods indicate the power of the molecular electronic descriptors Δ and η as tools for SAR studies.

In this work we initially investigate the ellipticines activity using EIM methodology. Then we use the PCA method in order to evaluate the relevance of EIM electronic parameters when compared to more traditional physicochemical descriptors. PCA results are corroborated by observing the overall similarity of molecules by HCA grouping.

PCA [48] is a multivariate exploratory method which has the aim to find linear transformations of original variables describing a set of samples (molecules in our case), into
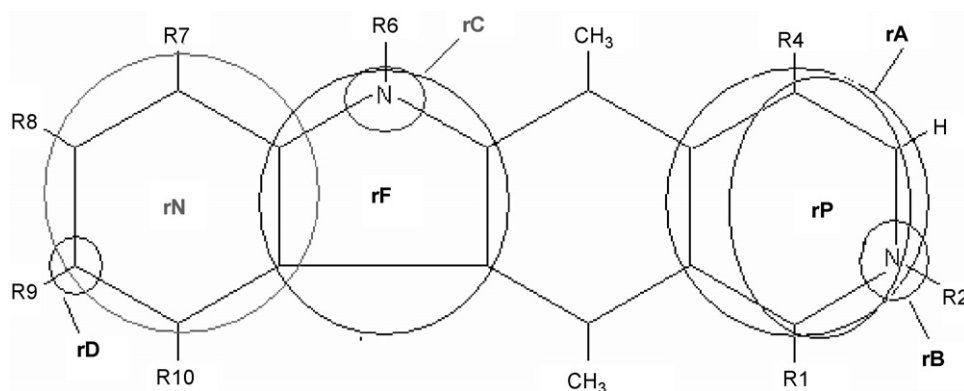

Fig. 2. Molecular regions selected for local electronic density of states calculations.

new uncorrelated variables able to separate them in distinct groups according to specific similarities. These new variables are called principal components (PC). The first PC is generated in a way that it maintains the statistic distribution of the samples and accounts for the largest portion of the total samples variance. The following PCs are constructed as orthogonal to the anterior (uncorrelated) and with the maximum possible of the remaining variance of the data set. A total of N PCs is obtained where N is the number of initial variables. Geometrically, a molecule is a point in a N-dimensional space of descriptors, and after the PCA transformation, the plotting of scores allows the observation of molecules grouping while the plotting of loadings identify the descriptors correlated with the pattern observed by the scores. One of the applications of PCA, during the investigation, is the identification of important descriptors and the elimination of variables which do not contribute to molecules separation according to biological activity.

HCA [48] is also exploratory and gives information of the global similarity between samples distributed in a multi-dimensional space. While PCA shows samples separation in a bidimensional space, HCA explores the N-dimensions simultaneously. The results are presented in a dendogram where Euclidean distances between samples are transformed in a similarity matrix and plotted in a range from zero to one. For details about dendograms see Refs. [24,47–49].

The EIM descriptors have been obtained with Chem2pac [49] program, and PCA and HCA studies were carried out using the program package Einsight [50].

In the next section we present and discuss the results of SAR investigation with the three above mentioned methods: EIM, PCA and HCA.

3. Results and discussion

We initially investigate the ellipticines activity using EIM. For the G1 group (molecules 1–20 and 32–36) we used this methodology to carry out a detailed analysis comparing the biological activity data (A, Table 1) and the electronic descriptors obtained for the seven regions above mentioned. The contribution of atoms of the region D to LUMO (CL(rD)) orbital showed the best correlation with the activity, correctly classifying 21 out of 25 molecules. If we take into account the difference in energy ΔH between the frontier orbital levels HOMO and HOMO-1 associated with CL(rD) the correct prediction of the activity increase to 23 out of 25 molecules.

The EIM classification with these two parameters is based on the simple rule that:

If $\Delta H \leq 0.85$ and $CL(rD) \geq 0.022$, the molecule will be inactive, otherwise the molecule will be active.

The ΔH and CL(rD) obtained with PM3 are presented in Table 2. From these data we observe that the molecules E15 and E18 are incorrectly classified as active. This corresponds to a global accuracy of 92% in the reproduction of experimental data. This is a very good result considering that only two descriptors are considered in the analysis. This level of reproduction of biologic data is similar to that obtained with the

Table 2
Descriptors DM, ΔH, CL(rD) and ηH(rB) obtained from PM3 calculations and elected in G1 group analysis

Molecule	DM (Debye)	ΔH (eV)	CL(rD)	ηH(rB)
E01	2.843	0.658	0.021	0.027
E02	2.884	0.808	0.023	0.035
E03	2.342	0.952	0.019	0.044
E04	2.608	0.669	0.034	0.067
E05	1.895	0.802	0.018	0.027
E06	1.984	1.131	0.025	0.076
E07	2.456	0.983	0.015	0.067
E08	1.862	1.087	0.020	0.098
E09	2.003	1.102	0.013	0.067
E10	2.868	0.967	0.008	0.059
E11	1.769	0.785	0.013	0.025
E12	2.840	0.636	0.020	0.024
E13	2.225	0.683	0.018	0.104
E14	2.852	0.692	0.029	0.029
E15	2.580	1.119	0.009	0.109
E16	2.631	0.980	0.015	0.063
E17	2.267	0.956	0.017	0.078
E18	2.222	0.979	0.017	0.073
E19	3.10	1.055	0.021	0.069
E20	2.618	1.004	0.017	0.065
E32	2.731	0.614	0.023	0.051
E33	1.728	0.402	0.028	0.068
E34	1.723	0.758	0.015	0.044
E35	0.979	1.094	0.020	0.086
E36	2.322	1.061	0.015	0.058

EIM methodology in previous studies of several families of molecules.

Using the rule derived from EIM methodology mentioned above we propose the classification of activity for the molecules of the G2 group. The descriptors calculated are presented in Table 3. We can observe that for the 15 untested molecules the EIM methodology proposes that 11 will be active.

In order to consider the weight of the electronic descriptors selected by EIM methodology we performed an independent PCA investigation with an initial set of 57 descriptors including all the parameters cited in the methodology. The parameters R, Po, M, V, HE, SE and log P have been disregarded after a

Table 3
Descriptors DM, ΔH, CL(rD) and ηH(rB) obtained from PM3

Molecule	DM (Debye)	ΔH (eV)	CL(rD)	ηH(rB)
E21	1.852	0.779	0.015	0.023
E22	2.012	0.795	0.008	0.009
E23	2.901	0.776	0.015	0.031
E24	3.134	0.669	0.022	0.031
E25	2.932	0.684	0.025	0.029
E26	3.038	0.603	0.022	0.038
E27	2.012	0.774	0.015	0.022
E28	2.918	0.922	0.005	0.056
E29	2.461	0.844	0.008	0.059
E30	1.102	1.085	0.023	0.069
E31	2.32	1.121	0.024	0.066
E37	3.751	0.763	0.017	0.047
E38	2.667	0.756	0.022	0.032
E39	2.433	0.548	0.020	0.081
E40	1.691	0.668	0.021	0.080

preliminary analysis due to their inability to differentiate molecules presenting distinct biological activities.

The best separation between active and inactive compounds was obtained using the following descriptors (see Table 2):

- the energy difference between HOMO and HOMO-1 (ΔH);
- the contributions to the LUMO orbital by local density of states (LDOS) over the molecular region D (CL(rD));
- the dipole moment values (DM);
- the differences of the HOMO and HOMO-1 contributions (CH and CH-1) to the local density of states (LDOS) over the molecular region B (ηH(rB)).

From the four descriptors selected, three are derived from EIM methodology. This choice was not biased during the PCA investigation and it emphasizes the importance of electronic parameters to biological activity identification of ellipticines. In Fig. 3, we show the scores of the first two principal components (PC1 and PC2) for the G1 group. The descriptors have been autoscaled before the analysis. The molecules are distributed into two distinct regions in the plane. The active group is on the right side and the inactive one on the left side, indicating that PC1 was mainly responsible for this characterization. The molecules E01 and E12 are incorrectly classified as inactive and the molecules E15 and E18 are incorrectly classified as active. This corresponds to 21 molecules out of the 25 correctly classified; which means an accuracy of 84%.

The plane PC1 × PC2 conserved 74% of the total variance of the original data, and the axis were written in terms of molecular descriptors according to the following equations:

$$PC1 = -0.30\,DM + 0.63\,\Delta H - 0.50\,CL(rD) + 0.52\,\eta H(rB)$$
(1)

$$PC2 = 0.76\,DM + 0.24\,\Delta H - 0.50\,CL(rD) - 0.33\,\eta H(rB)$$
(2)

Eq. (1) indicates no significant differences in the contributions of the descriptors to PC1 while Eq. (2) indicates that the dipole

Fig. 4. Plotting of loading for the four descriptors related to the ellipticine (G1 group) classification.

moment contribution to PC2 is greater than the other ones. In a general way the descriptors selected by PCA are in accordance with previous investigations of ellipticines. The DM exhibits correlation to the activity ellipticines [28–30] and MEP over D region indicates it as an active site [30]. At the same time experimental investigations [5] identify the D region as relevant to the biological activity of ellipticines. On the other hand, ΔH and ηH(rB) have not yet been considered in previous ellipticines investigation and are new parameters that we propose as appropriate to further investigation.

In Fig. 4, we show the loading vectors of the four physicochemical descriptors that have oriented the molecules separation. The plot indicates that the descriptors CL(rD)/DM and $\Delta H/\eta H$(rB) has the same effect of pulling the active and inactive molecules towards distinct regions of the graphic.
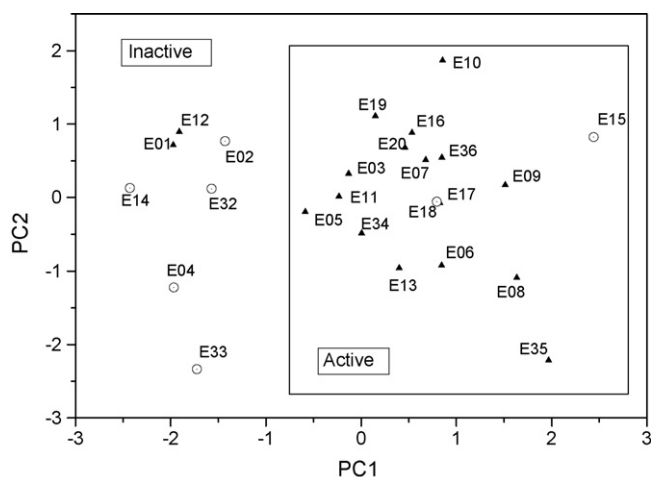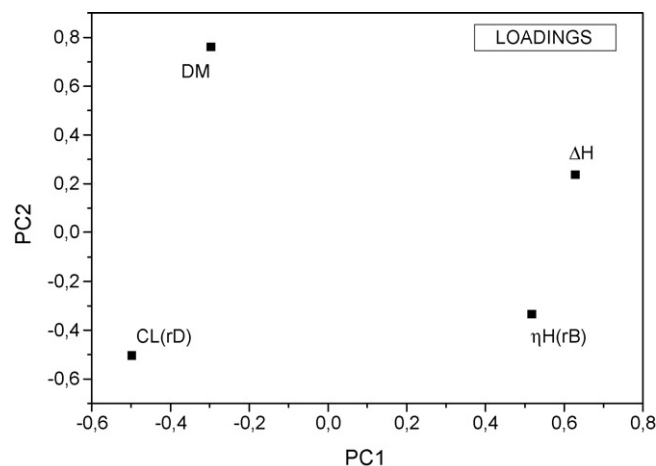

Fig. 3. Plotting of score data with the separation of the 25 compounds of G1 group into two subgroups as active and inactive in the PC1 × PC2 space. Triangle and ring symbols represent active and inactive molecules, respectively.
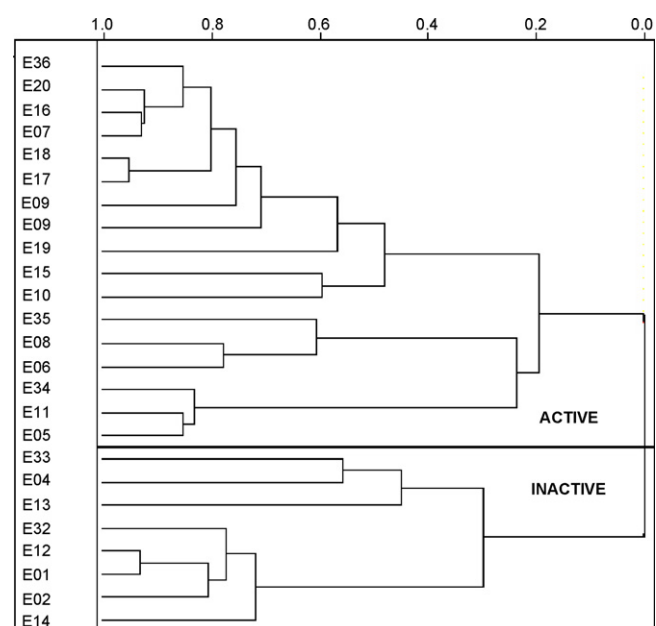

Fig. 5. Dendogram of the hierarchical cluster of the G1 group. The darker line indicates the active and inactive clusters separation.

These PCA results are corroborated by the hierarchical clustering diagram plot for the G1 group presented in Fig. 5. We can see that with the use of four descriptors of PCA calculations two clusters are formed. They are separated by the horizontal line between the E05 and E33 molecules. Most of the active molecules are in the upper half group while the inactive populates the lower half one. These two clusters have zero similarity, which demonstrates that active and inactive molecules are well separated in the four-dimensional PCA space. In this analysis five molecules are incorrectly classified using HCA: E01, E12, E15, E18 and E13.

We now consider the results obtained with PCA for G1 group in order to speculate about the activity of G2 molecules. As PCA is not a classificatory method, the results and the information that can be obtained for new, untested molecules must be interpreted as an indication of the affinity of new ellipticines with groups already formed in previous studies.

We, then, applied the PCA method to G1 and G2 molecules together using exactly the same four descriptors selected for G1 group. The PCA score graph of the first two principal components (PC1 and PC2) is illustrated in Fig. 6. Again, the molecules are distributed into two distinct regions in the figure indicating that PC1 was determinant for this characterization.

Autoscaling the new set of data we have generated new equations for the PCs axis. The two principal components (PC1 and PC2) are given by Eqs. (3) and (4) (see Table 3):

$$PC1 = -0.44\,DM + 0.64\,\Delta H - 0.30\,CL(rD) + 0.55\,\eta H(rB) \tag{3}$$

$$PC2 = -0.40\,DM - 0.22\,\Delta H + 0.81\,CL(rD) + 0.37\,\eta H(rB) \tag{4}$$

Using the distribution showed in Fig. 6 we can propose a classification for the G2 molecules as active or inactive. We have found that 9 out of 15 molecules are classified as active. The PCA and EIM predictions disagree for molecules E23 and E37: they are described as active by EIM but the PCA has classified them as inactive. The loadings graph of the new
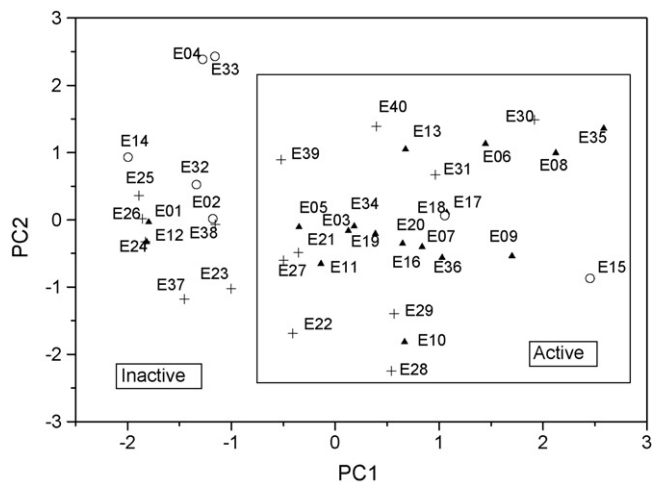

Fig. 6. Plotting of score data with the separation of the global set of ellipticines (G1 + G2) into two subgroups as active and inactive, in the PC1 × PC2 space. Triangle, ring and cross symbols represent active, inactive and nonstested molecules, respectively.

investigation present a result similar to the previous one (Fig. 4).

We also evaluate the global similarity of the total group of 40 molecules with the HCA method. Fig. 7 shows the dendogram for the entire set of molecules (G1 + G2) separated into two distinct clusters of active and inactive molecules. Compared to EIM and PCA results (Table 4) only the E39 and E40 molecules are now classified as inactive. The HCA results are in good agreement with the previous investigation of Ref. [30] (Table 4), with only four molecules (E28, E29, E39 and E40) now included in a different class of activity.

In Table 4, we compare our prediction of the antitumor activity for the group of 15 untested molecules: EIM indicates
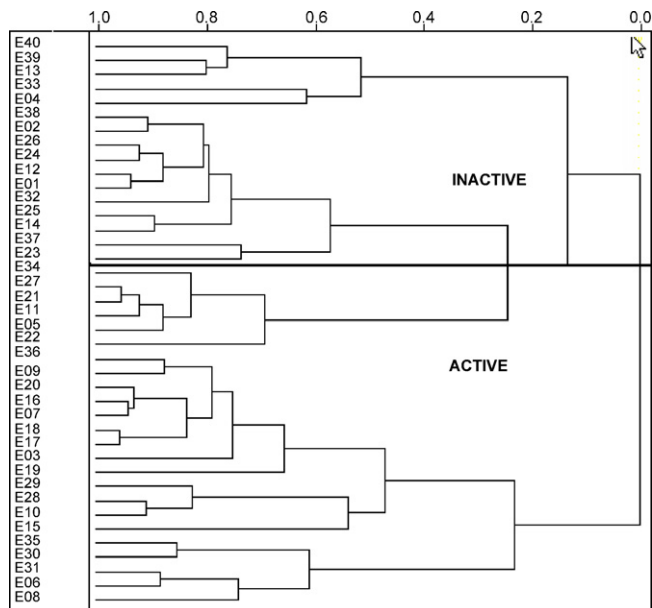
Table 4
Proposed activity (PA) for molecules of G2 group by EIM, PCA and HCA analysis compared with Ref. [30]

Molecule	PA EIM	PA PCA	PA HCA	PA Ref. [30]
E21	A	A	A	A
E22	A	A	A	A
E23	A	I	I	I
E24	I	I	I	I
E25	I	I	I	I
E26	I	I	I	I
E27	A	A	A	A
E28	A	A	A	I
E29	A	A	A	I
E30	A	A	A	A
E31	A	A	A	A
E37	A	I	I	I
E38	I	I	I	I
E39	A	A	I	A
E40	A	A	I	A

A and I refer to active and inactive, respectively.

Fig. 7. Dendogram of the hierarchical distribution of the entire set of molecules (G1 + G2 groups). The darker line indicates the active and inactive clusters separation.

that 11 will be active, PCA indicates that nine will be active while HCA indicates that seven will be active. Combining these results, and considering as a condition for activity that at least two methods classify the molecule as active, we have that nine molecules should be active and the remaining ones should be inactive. This classification agrees with the results presented in Ref. [30] for all the molecules.

4. Conclusions

Ellipticine and many of its derivatives are a group of drugs which exhibit significant antitumor properties both *in vitro* and *in vivo* and which are used in the chemotherapy of leukemia and other different forms of cancer [51]. In this work we investigate a set of 40 ellipticine derivatives and analogues. Three different pattern-recognition methodologies have been applied in the analysis with dozens of electronic, stereochemical and physicochemical descriptors.

We considered the molecules divided in two groups of tested (G1) and nontested (G2) molecules. Taking into account simple rules, our results show that the electronic descriptors ΔH (energy difference between the frontier orbitals HOMO and HOMO-1), CL(rD) and CH(rB) (which are related to the distribution of LUMO and the HOMO orbitals over isolated atoms of the exposed rings of the molecular skeleton) correctly identifies ellipticines biological activity in G1 with 92% of accuracy.

The importance of these descriptors has been reinforced by the multivariate analysis of PCA and HCA. Three out of four descriptors selected by PCA (ΔH, CL(rD) and CH(rB)) are derived from EIM methodology and the remaining descriptor (DM) has been selected in our previous work.

The three different methods reproduce experimental results with a high level of accuracy. EIM correctly classifies 92% of the molecules, while PCA and HCA achieve 84% and 80% of correct results, respectively. These results indicate that molecular properties are the most relevant ones in order to distinguish active from inactive molecules. A similar result was reported by Begnini and Passerini concerning the carcinogenic activity of aromatic amines which depends mainly on electronic and steric properties [52]. On the other hand, the gradation of potency of the active molecules could depend on their interactions with the biological medium.

The rules and pattern of classification of activity obtained for the 25 molecules of the G1 group have been applied to propose the activity of the untested molecules of G2. Combining the results of the three methods we propose that from 15 molecules 9 should be active in agreement with previous results [30]. Although PCA and HCA are not classificatory methods, they are useful to provide an independent analysis of the EIM results for the untested molecules. Furthermore, preliminary results for the untested molecules, obtained using the non-algorithmic analysis of Artificial Neural Networks, support the conclusions found in the present work.

The set of electronic parameters selected as relevant to the antitumor activity from our analysis might be useful for future investigations. Other studies in this way are in progress.

Acknowledgments

References

[1] S. Goodwin, A.F. Smith, E.C. Horning, Alkaloids of *Ochrosia-Elliptica Labill*, J. Am. Chem. Soc. 81 (1959) 1903–1908; R.B. Woodward, G.A. Iacobucci, F.A. Hochstein, The synthesis of ellipticine, J. Am. Chem. Soc. 81 (1959) 4434–4435.

[2] G.W. Gribble, Syntheis and antitumor activity of elipticine alkaloids and related compounds, in: A. Brossi (Ed.), The Alkaloids, vol. 39, Academic Press, New York, 1990, pp. 239–343.

[3] P. Sizun, C. Auclair, E. Lescot, C. Paoletti, B. Perly, S. Fermandjian, Stacking and edge-to-edge associations of antitumoral ellipticine derivatives are controlled in solution by interactions involving their nitrogen sites, Biopolymers 27 (1988) 1085–1096.

[4] D. Reha, M. Kabelác, F. Ryjácek, J. Sponer, J.E. Sponer, M. Elstner, S. Suhai, P. Hobza, Intercalators. 1. Nature of stacking interactions between intercalators (ethidium, daunomycin, ellipticine, and 4,6-diaminide-2-phenylindole) and DNA base pairs. Ab initio quantum chemical, density functional theory, and empirical potential study, J. Am. Chem. Soc. 124 (2002) 3366–3376.

[5] C. Auclair, Multimodal action of antitumor agents on DNA: the ellipticines series, Arch. Biochem. Biophys. 259 (1987) 1–14.

[6] O. Mauffret, B. Rene, O. Convert, M. Monnot, E. Lescot, S. Fernandjian, Drug–DNA interactions: spectroscopic and footprinting studies of site and sequence specificity of ellipticinium, Biopolymers 31 (1991) 1325–1341.

[7] M.A. Schwaller, G. Dodin, J. Aubard, Thermodynamics of drug–DNA interactions: entropy-driven intercalation and enthalpy-driven outside binding in the ellipticine series, Biopolymers 31 (1991) 519–527.

[8] A. Adenier, J. Aubard, M.A. Schwaller, Thermodynamics and kinetics of aggregation processes in aqueous-media of ellipticine derivatives—the alkyl oxazolopyridocarbazole series, J. Phys. Chem. 96 (1992) 8785–8791.

[9] J. Aubard, M.A. Schwaller, J. Patigny, J.P. Marsault, G. Lévi, Surface-enhanced raman-spectroscopy of ellipticine, 2-*N*-methylellipticinium and their complexes with DNA, J. Raman Spectrosc. 23 (1992) 373–377.

[10] K.W. Kohn, M.J. Waring, D. Glaubiger, C. Friedman, Intercalative binding of ellipticine to DNA, Cancer Res. 35 (1975) 71–76.

[11] M. Stiborová, M. Stiborová-Rupertová, L. Borjek-Dohalská, M. Wiessler, E. Frei, Rat microsomes activating the anticancer drug ellipticine to species covalently binding to deoxyguanosine in DNA are a suitable model mimicking ellipticine bioactivation in humans, Chem. Res. Toxicol. 16 (2003) 38–47.

[12] M. Stiborová, C.A. Bieler, M. Wiessler, E. Frei, The anticancer agent ellipticine on activation by cytochrome P450 forms covalent DNA adducts, Biochem. Pharmacol. 62 (2001) 1675–1684.

[13] J.B. Le Pecq, N. Dat-Xuong, C. Gosse, C.A. Paoletti, New antitumoral agent – 9-hydroxyellipticine – possibility of a rational design of anti-cancerous drugs in series of DNA intercalating drugs, Proc. Natl. Acad. Sci. U.S.A. 71 (1974) 5078–5082.

[14] C. Paoletti, S. Cros, N. Dat-Xuong, P. Lecointe, A. Moisand, Comparative cytotoxic and anti-tumoral effects of ellipticine derivatives on mouse-L-1210 leukemia, Chem. Biol. Interact. 25 (1979) 45–58.

[15] H. Chabane, C. Lamazzi, V. Thiéry, G. Guillaumetb, T. Bessona, Synthesis of novel 2-cyanothiazolocarbazoles analogues of ellipticine, Tetrahedron Lett. 43 (2002) 2483–2486.

[16] H.A. Tran-Thi, T. Nguyen-Thi, S. Michel, F.K.O.C.H. Tillequin, M. Pfeiffer, B. Pierré, A. Trinh-Van-Dufat, Synthesis and cytotoxic activity

of pyranocarbazole analogues of ellipticine and acronycine, Chem. Pharm. Bull. 52 (2004) 540–545.

[17] Y. Ergun, S. Patir, G. Okay, A novel synthesis towards ellipticine and its derivatives. Synthesis of a new precursor compound, Synth. Commun. 34 (2004) 435–442.

[18] M. Ishikura, A. Hino, T. Yaginuma, I. Ágata, N. Katagiri, A novel entry to pyrido[4,3-b]carbazoles: an efficient synthesis of ellipticine, Tetrahedron 56 (2000) 193–207.

[19] M. Diaz, A. Cobas, E. Guitian, L. Castelo, Synthesis of ellipticine by hetaryne cycloadditions—control of regioselectivity, Eur. J. Org. Chem. 23 (2001) 4543–4549.

[20] M. Ishikura, A. Hino, N. Katagiri, An efficient total synthesis of ellipticine, Heterocycles 53 (2000) 11–14.

[21] B. Diop, P. Toure, M.T. Sow, M. Toure, M.L. Halliez, J.P. Castaigne, J.M. Mondesir, R. De Jaeger, Med. Afr. Noire 31 (1984) 107–110.

[22] L.K. Dalton, S. Demerac, B.C. Elmes, J.W. Loder, J.M. Swan, J. Teitei, Synthesis of tumour-inhibitory alkaloids ellipticine 9-methoxyellipticine and related pyrido[4,3-B]carbazoles, Aust. J. Chem. 20 (1967) 2715–2727.

[23] R. JasztoldHoworko, et al., Synthesis and evaluation of 9-hydroxy-5-methyl- and 5,6-dimethyl)-6h-pyrido[4,3-B]carbazole-1-*N*-[(dialkylamino)alkyl]carboxamides, a new promising series of antitumor olivacine derivatives, J. Med. Chem. 37 (1994) 2445–2452.

[24] W.K. Anderson, A. Gopalsamy, P.S. Reddy, Design, synthesis, and study of 9-substituted ellipticine and 2-methylellipticinium analogs as potential CNS-selective antitumor agents, J. Med. Chem. 37 (1994) 1955–1963;
J. Jurayj, R.D. Haugwitz, R.K. Varma, K.D. Paull, J.F. Barret, M. Cushman, Design and synthesis of ellipticinium salts and 1,2-dihydroellipticines with high selectivities against human CNS cancers in-vitro, J. Med. Chem. 37 (1994) 2190–2197.

[25] G. Mathe, P. Pontiggia, C. Bourut, E. Chenu, S. Orbach-Arbouys, *In vivo* eradication of friend-virus as an experimental HIV-model, by combination of zidovudine, acriflavine and an ellipticine analog—possible application to the treatment of human pre-AIDS, Biom. Pharmacother. 48 (1994) 51–53.

[26] J.C. Ruckdeschel, S.P. Modi, W. El-Hamouly, E. Portuese, S. Archer, *N*-Methylcarbamate derivatives of ellipticines and olivacine with cytotoxic activity against four human lung cancer lines, J. Med. Chem. 35 (1992) 4854–4857.

[27] K.W. Kohn, W.E. Ross, D. Glaubinger, in: F.E. Hahn (Ed.), Antibiotics, vol. 2, Springer-Verlag, Berlin, 1979, p. 195.

[28] S.O. Dantas, F.C. Lavarda, B. Laks, D.S. Galvão, An investigation of the electronic structure of the antitumor drug ellipticine and its derivatives. Part I. Geometrical AM1 Study, J. Mol. Struct. (Theochem.) 253 (1992) 319–332.

[29] P.M.V.B. Barone, S.O. Dantas, D.S. Galvão, A semi-empirical study on the electronic structure of ellipticines, J. Mol. Struct. (Theochem.) 465 (1999) 219–229.

[30] S.F. Braga, L.C. de Melo, P.M.V.B. Barone, Semiempirical study on the electronic structure of antitumor drugs ellipticines, olivacines and iso-ellipticines, J. Mol. Struct. (Theochem.) 710 (2004) 51–59.

[31] P.M.V.B. Barone, A. Camilo Jr., D.S. Galvão, Theoretical approach to identify carcinogenic activity of polycyclic aromatic hydrocarbons, Phys. Rev. Lett. 77 (1996) 1186–1189.

[32] R.S. Braga, P.M.V.B. Barone, D.S. Galvão, Identifying carcinogenic activity of methylated polycyclic aromatic hydrocarbons (PAHs), J. Mol. Struct. (Theochem.) 464 (1999) 257–266.

[33] D. Villemin, D. Cherqaoui, A. Mesbah, Predicting carcinogenicity of polycyclic aromatic-hydrocarbons from back-propagation neural-network, J. Chem. Inf. Comput. Sci. 34 (1994) 1288–1293.

[34] E.L. Cavaliere, E.G. Rogan, R.W. Roth, R.K. Saugier, A. Hakam, The relationship between ionization-potential and horseradish-peroxidase hydrogen-peroxide batalyzed binding of aromatic-hydrocarbons to DNA, Chem. Biol. Interact. 47 (1983) 87–109.

[35] CACHE, version 5.0, Fujitsu Limited, Chiba City, Japan, 2001.

[36] M.C. Zerner, Semiempirical molecular orbital methods, in: K.B. Lipkowitz, D.B. Boyd (Eds.), Reviews in Computational Chemistry, vol. 2, VCH Publishers, New York, 1991, pp. 313–365.

[37] P.M.V.B. Barone, R.S. Braga, A. Camilo, D.S. Galvão, Electronic indices from semiempirical calculations to identifying carcinogenic activity of polycyclic aromatic hydrocarbons (PAHs), J. Mol. Struct. (Theochem.) 505 (2000) 55–66.

[38] R. Vendrame, R.S. Braga, Y. Takahata, D.S. Galvão, Structure–activity relationship studies of carcinogenic activity of polycyclic aromatic hydrocarbons using calculated molecular descriptors with principal component analysis and neural network methods, J. Chem. Inf. Comput. Sci. 39 (1999) 1094–1104.

[39] L.L.D. Santo, D.S. Galvão, Structure–activity study of indolequinones bioreductive alkylating agents, J. Mol. Struct. (Theochem.) 464 (1999) 273–279.

[40] S.F. Braga, D.S. Galvão, A structure–activity study of taxol, taxotere and derivatives using the electronic indices methodology (EIM), J. Chem. Inf. Comput. Sci. 43 (2003) 699–706.

[41] R. Vendrame, R.S. Braga, D.S. Galvão, Structure–activity relationship (SAR) studies of the tripos benchmark steroids, J. Mol. Struct. (Theochem.) 619 (2002) 195–205.

[42] R.S. Braga, R. Vendrame, D.S. Galvão, Structure–activity relationship studies of substituted 17 alpha-acetoxyprogesterone hormones, J. Chem. Inf. Comput. Sci. 40 (2000) 1377–1385.

[43] M. Cyrillo, D.S. Galvão, Structure–activity relationship study of some inhibitors of HIV-1 integrase, J. Mol. Struct. (Theochem.) 464 (1999) 267–272.

[44] N.W. Ashcroft, N.D. Mermim, Solid State Physics, Saunders College, Philadelphia, 1976.

[45] G.G. Nys, R.F. Rekker, Concept of hydrophobic fragmental constants (*F*-values). 2. Extension of its applicability to calculation of lipophilicities of aromatic and heteroaromatic structures, Eur. J. Med. Chem. Chim. Ther. 9 (1974) 361–375.

[46] S.F. Braga, D.S. Galvão, Benzo[*c*]quinolizin-3-ones theoretical investigation: SAR analysis and application to nontested compounds, J. Chem. Inf. Comput. Sci. 44 (2004) 1987–1997.

[47] T. Naes, P. Baardseth, H. Helgesen, T. Isakson, Multivariate techniques in the analysis of meat quality, Meat Sci. 43 (1996) s135–s149.

[48] D.L. Massart, B.G.M. Vandeginste, S.N. Deming, Y. Michotte, L. Kaufman, Chemometrics: A Textbook, Elsevier, Amsterdam, 2003.

[49] Cyrillo, M., Galvão, D.S., Chem2Pac: a computational chemistry integrator for Windows. EPA Newslett. 67 (1999) 31–38.

[50] EinSight, version 3.0, Infometrix, Inc., Seattle, WA, 1991.

[51] L. Larue, C. Rivalle, G. Muzard, C. Paoletti, E. Bisagni, J. Paoletti, A new series of ellipticine derivatives (1-(alkylamino)-9-methoxyellipticine)—synthesis, DNA-binding and biological properties, J. Med. Chem. 31 (1988) 1951–1956.

[52] R. Benigni, L. Passerini, Carcinogenicity of the aromatic amines: from structure–activity relationships to mechanisms of action and risk assessment, Mutat. Res. 511 (2002) 191–206.