Available online at www.sciencedirect.com
SCIENCE DIRECT®
European Journal of Medicinal Chemistry 40 (2005) 329–338
EUROPEAN JOURNAL OF MEDICINAL CHEMISTRY
www.elsevier.com/locate/ejmech

Original article

A structure–activity relationship study of quinone compounds with trypanocidal activity

F.A. Molfetta [a], A.T. Bruni [b], K.M. Honório [a], A.B.F. da Silva [a,*]

[a] *Departamento de Química e Física Molecular, Instituto de Química de São Carlos, Universidade de São Paulo, CP 780, 13560-970 São Carlos, SP, Brazil*
[b] *Departamento de Física, IBILCE, Instituto de Biociências, Letras e Ciências Exatas, Universidade Estadual Paulista, 15054-000 São Jose do Rio Preto, SP, Brazil*

Received 7 May 2004; received in revised form 26 October 2004; accepted 27 October 2004

Available online 01 February 2005

Abstract

A set of 25 quinone compounds with anti-trypanocidal activity was studied by using the density functional theory (DFT) method in order to calculate atomic and molecular properties to be correlated with the biological activity. The chemometric methods principal component analysis (PCA), hierarchical cluster analysis (HCA), stepwise discriminant analysis (SDA), *K*th nearest neighbor (KNN) and soft independent modeling of class analogy (SIMCA) were used to obtain possible relationships between the calculated descriptors and the biological activity studied and to predict the anti-trypanocidal activity of new quinone compounds from a prediction set. Four descriptors were responsible for the separation between the active and inactive compounds: T_5 (torsion angle), QTS1 (sum of absolute values of the atomic charges), VOLS2 (volume of the substituent at region B) and HOMO-1 (energy of the molecular orbital below HOMO). These descriptors give information on the kind of interaction that occurs between the compounds and the biological receptor. The prediction study was done with a set of three new compounds by using the PCA, HCA, SDA, KNN and SIMCA methods and two of them were predicted as active against the *Trypanosoma cruzi*.

Keywords: Quinone; Anti-trypanocidal activity; DFT; Pattern recognition; Structure–activity relationship; Chemometrics

1. Introduction

Chagas disease, or American trypanosomiasis, is endemic in Central and South America and an estimated 16–18 million persons are infected with *Trypanosoma cruzi* (Kinetoplastida: Trypanosomatidae). The *T. cruzi* is the causative agent of Chagas disease and 100 million people (roughly one-quarter of the population) are at risk of infection. Despite decreasing rates of prevalence and incidence of *T. cruzi* infection, Chagas disease remains a serious obstacle to health and economic development in Latin America, especially for the poor rural population [1]. The drug Benznidazole has been demonstrated to have an efficacy of 56% in the treatment of acute and indeterminate cases of the disease. However, there is presently no available drug recommended for patients with evidence of cardiomyopathy or one of megasyndromes (mega colon or mega esophagus). Thus, new drugs with improved efficacy are needed for the treatment of this disease [2].

Quinone compounds are common to numerous natural products and they are associated with anticancer, antibacterial, antimalarial and fungicide activities. In most cases, the biological activity is related to the ability of quinones to accept one and/or two electrons to form the corresponding radical anion or dianion species as well as the acid–base properties of the compounds. The variable capacity of quinone compounds to accept electrons is due to the electron-attracting (or donating) substituents at the quinone moiety which modulate the redox properties responsible for the resulting oxidative stress [3]. The cytostatic and antiparasitic activities of some quinones can also emerge due to their ability to act as

* Corresponding author. Tel.: +55 16 33 739 964; fax: +55 16 33 739 975.
E-mail address: alberico@iqsc.usp.br (A.B.F. da Silva).

0223-5234/$ - see front matter © 2005 Elsevier SAS. All rights reserved.
doi:10.1016/j.ejmech.2004.10.009

potent inhibitors of electron transport [4], as uncouples of oxidative phosphorylation [5], as intercalating agents in DNA double helix [6] or as bioreductive alkylating agents of biomolecules [7].

Molecular modeling depends on the type of available information. Two different situations can be evaluated when a structure–activity relationship study is performed: the active-site of the receptor is known (from X-rays or NMR analysis) or unknown. For the first case, molecular modeling is applied for designing a molecule with complementary structure to the active-site. When the active-site is unknown, SAR or QSAR techniques can be applied to a series of similar compounds with known biological activity previously determinated.

The main goal of this work is to investigate the SAR of compounds listed in Fig. 1 using atomic and molecular descriptors (electronic, steric, hydrophobic and topological). For that, we made use of the density functional theory (DFT) method to calculate atomic and molecular descriptors of 25 quinone compounds (training set) reported in the literature as potent and selective trypanocidal agents [8,9]. The calculated descriptors were selected so that some steric, electronic, hydrophobic and topological characteristics of these compounds could be taken into account since each one of them can contribute to the biological activity and give information about the interactions between the compounds and the biological receptor. Chemometric techniques such as principal component analysis (PCA), hierarchical cluster analysis (HCA), stepwise discriminant analysis (SDA), *K*th nearest neighbor (KNN) and soft independent modeling of class analogy (SIMCA) were employed to analyze the data set and to obtain the relationship between the atomic and molecular descriptors and the biological activity. The results obtained with PCA, HCA, SDA, KNN and SIMCA were tested in three new quinone compounds (prediction set) [10,11] with the aim to make the activity prediction of these new compounds and also validate our models.

2. Methods

The central structure, numbering and the chemical structure of the 25 quinone compounds studied in this work are presented in Fig. 1. The compounds shown in Fig. 1 consist of 15 active molecules and 10 inactive molecules against *T. cruzi*.

The geometry optimizations of the quinone compounds were performed with the initial structures (see Fig. 1) by using the DFT/B3LYP method [11], which uses the Becke's theory [12] and the Lee–Yang–Parr correlation function [13]. The choice for the DFT method was made because recent studies have demonstrated that the DFT/B3LYP method leads to excellent results for the analysis of geometries and energies [14,15].

The molecular properties (variables or descriptors) of the 25 quinone compounds were calculated using the

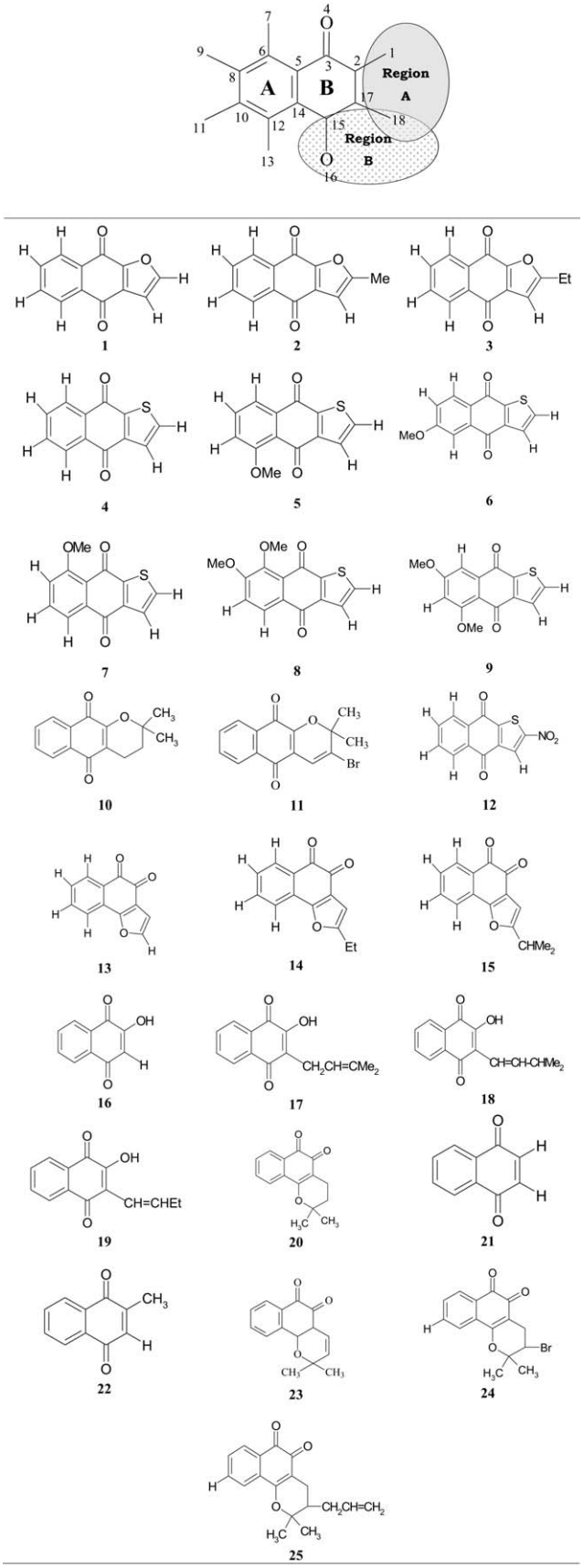

Fig. 1. The central chemical structure, numbering used and chemical structure of the 25 quinone compounds studied.

DFT/B3LYP method with the 6-31G* basis set, as implemented in the GAUSSIAN 98 computational package [16], from the minimum energy conformation obtained for each compound. The following molecular descriptors were calculated: total energy, energy of the highest occupied molecular orbital (HOMO), energy of the molecular orbital below HOMO (HOMO-1), energy of the lowest unoccupied molecular orbital (LUMO), differences of some molecular orbital energies, bond orders over all bonds that comprise the basic skeleton of the quinone compounds studied, molecular hardness, molecular softness, dipole moment, atomic charge Q_i, $i = 1, 2, ..., 18$ (see Fig. 1), sum of the atomic charges of the substituents at regions A and B, molecular volume, octanol/water partition coefficient (log P), torsion angle formed by 1–4 atoms and a large variety of topological descriptors. In fact, we calculated about 223 descriptors to be correlated with the trypanocidal activity of the quinone compounds under study. The atomic charges of the quinones were calculated with the CHELPG option in the GAUSSIAN 98 program. The CHELPG option allow us to evaluate the electrostatic potential of a molecule from the calculation of a set of punctual atomic charges so that it represents the possible best quantum molecular potential for a set of points defined around the molecule [17]. The topological descriptors were evaluated from the WHIM/3D-QSAR molecular package [18]. The WHIM descriptors contain information about the whole 3D molecular structure in terms of size, shape, symmetry and atom distribution [19].

After the calculation of atomic and molecular descriptors, the Fisher's weights of these descriptors were obtained and the more significant descriptors were selected, i.e. those that had greatest Fisher weights were considered with a high ability in the discrimination (separation) between active and inactive compounds. The pattern recognition methods such as PCA, HCA, SDA, KNN and SIMCA were employed to study the relationship between the selected descriptors and the biological activity. Before employing the pattern recognition methods PCA, HCA, SDA, KNN and SIMCA all calculated variables (descriptors) were autoscaled so that they could be compared to each other on the same scale.

3. Results and discussion

3.1. Principal component analysis (PCA)

The PCA was first described by Pearson in 1901 and by Hotelling in 1933. In the past years, the use of PCA has increased and now it is often applied in the field of chemometrics [20]. It aims to group these correlated variables, generating a new set of variables called principal components (PCs). These PCs are built as linear combination of the original variables and have the important property of being completely uncorrelated. The first new axis, PC1, is chosen in the direction that maximizes the variance; the second axis must be chosen orthogonal to the first one and in the direction to describe as much variance left as possible and so on [21].

Table 1
Values of the four most important properties (variables) that classify the 25 quinone compounds studied

Compound	T_5 (°)	QTS1	VOLS2 (Å^3)	HOMO-1 (eV)
1	106.53	0.021	243.56	–0.3545
2	106.47	0.002	243.09	–0.3522
3	106.46	0.014	242.91	–0.3520
4	112.93	–0.018	241.33	–0.3541
5	112.84	–0.049	241.43	–0.3449
6	112.96	–0.049	242.02	–0.3501
7	113.17	–0.046	242.46	–0.3404
8	113.07	–0.074	242.31	–0.3421
9	112.80	–0.075	241.44	–0.3449
10	120.80	–0.001	248.43	–0.3488
11	118.35	–0.056	354.71	–0.3557
12	112.84	–0.097	241.98	–0.3711
13	132.55	–0.097	261.14	–0.3614
14	132.66	–0.161	373.62	–0.3591
15	132.60	–0.149	418.22	–0.3588
16	121.75	–0.090	200.43	–0.3570
17	123.09	–0.018	450.33	–0.3495
18	123.80	–0.034	454.67	–0.3525
19	123.82	–0.025	405.35	–0.3526
20	118.19	–0.130	420.39	–0.3528
21	122.40	–0.093	187.59	–0.3600
22	121.49	–0.053	200.99	–0.3566
23	123.24	–0.102	410.63	–0.3570
24	123.21	–0.048	466.21	–0.3630
25	123.31	–0.180	540.66	–0.3562

The initial data matrix, represented by **X**, is decomposed into two matrices, **T** and **P**, where

$$\mathbf{X} = \mathbf{T}\mathbf{P}^{\mathbf{T}} \quad (1)$$

In Eq. (1), **T** is known as the scores matrix and represents the position of the samples in the new coordinate system. The matrix **P** in Eq. (1) is known as the loadings matrix and describes how the new axes, i.e. the PCs, are built from the original variables. The samples are mapped through the scores and the variables by the loadings in the new low dimensional vector space defined by the PCs [21].

After several attempt to obtain a good classification for the compounds, the best separation was obtained with four variables (see Table 1): T_5 (torsion angle formed by atoms 1, 2, 17 and 18), QTS1 (sum of the atomic charges of C1, C2, C17, C18 and the substituents at region A), VOLS2 (total volume formed by the atoms C15, O16, C17, C18 and the substituents at region B) and HOMO-1 (energy of the molecular orbital below HOMO). This suggests that the other variables are not relevant for classifying the compounds studied according to their anti-trypanocidal activity.

The first three PCs explain 91.1% of the total variance of the data as follows: PC_1 = 56.2%, PC_2 = 20.8% and PC_3 = 14.1%. A number of score plots were examined and the most informative one is presented in Fig. 2 (which shows the first PC against the second component). This projection keeps 77.0% of the total variance of the original data and can be expected to provide a reasonably accurate representation

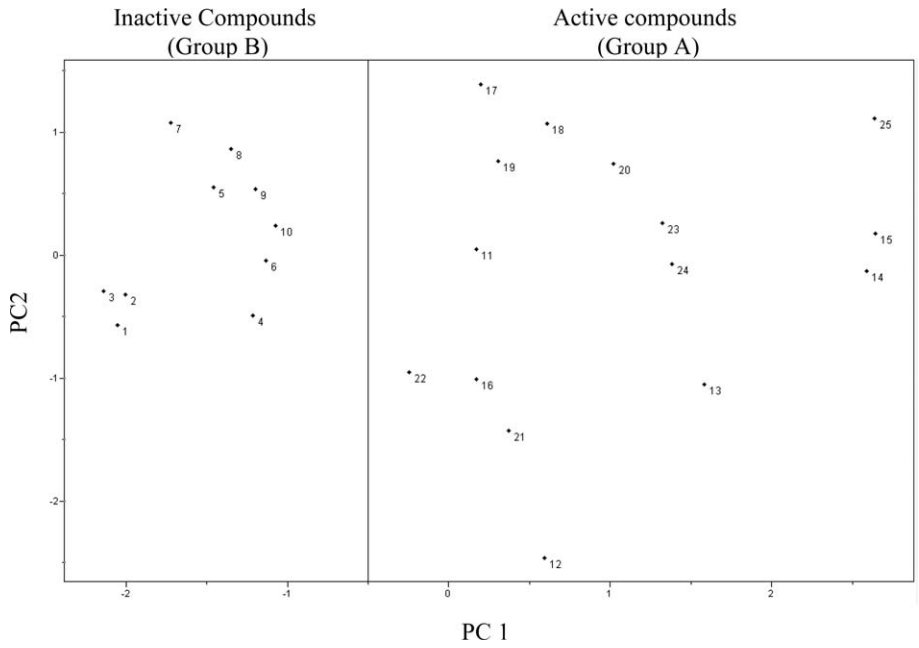

Fig. 2. Plot of the first two score vectors (PC_1 and PC_2) for the 25 quinone compounds with anti-trypanocidal activity. The PC analysis leads to a separation into two groups: active and inactive compounds.

of the higher order space. Table 2 shows the loading vectors for PC_1 and PC_2.

The plot of the score vectors for the first two PCs ($PC_1 \times PC_2$) is shown in Fig. 2. From Fig. 2, it can be seen that the quinone compounds studied are separated into two groups: A (active compounds, **11**–**25**) and B (inactive compounds, **1**–**10**). Also from Fig. 2, we can see that PC_1 alone is responsible for the separation between the active and inactive compounds.

Fig. 3 displays the plot of the loading vectors for the first two PCs (PC_1 and PC_2) and Table 2 shows the loading values for the selected variables in PC_1 and PC_2. According to Fig. 3, the HOMO-1 and QTS1 descriptors are responsible for describing the inactive compounds and the T_5 and VOLS2 descriptors are responsible for describing the active ones. According to Table 2, PC_1 can be expressed through the following equation:

$$PC_1 = 0.579\,[T_5] - 0.539\,[QTS1] + 0.455\,[VOLS2] - 0.406\,[HOMO\text{-}1] \quad (2)$$

From Eq. (2), we can see that the active molecules ($PC_1 > 0$) can be obtained when we have higher values for T_5 and VOLS2 combined with negative values for the variables HOMO-1 and QTS1. In this way, some conclusions on the most important variables can be drawn for the active molecules:

Table 2
The loading values for the first two PCs

Variable	PC_1	PC_2
T_5	0.579	0.073
QTS1	–0.539	0.019
VOLS2	0.455	0.623
HOMO-1	–0.406	0.778

1. T_5: from Table 1 we can see that the variable T_5 has higher values for the active compounds than the inactive ones, and this indicates that the active compounds need to have a suitable conformation, determined by the torsion angle formed by the atoms C1, C2, C3 and C4 (T_5), so that they may effectively interact with the biological receptor.
2. QTS1: regarding the sum of the charges of the atoms C1, C2, C17, C18 and the substituents at region A (QTS1), we can see that the active compounds need to have electron-acceptor atoms at region A, as $PC_1 > 0$ for this class of compounds.
3. VOLS2: according to Eq. (2), the active compounds need to have high values for VOLS2 (the volume formed by the atoms C15, O16, C17, C18 and the substituents at region B). The variable VOLS2 can be related to the fitting between the compound and the receptor.
4. HOMO-1: the energy of the frontier orbitals are important properties in several chemical and pharmacological processes, and the reason for this is the fact that these properties give information on the electron-donating and electron-accepting character of a compound, i.e. on the formation of a charge transfer complex. From Table 1 and Fig. 4 we can see that the energy of HOMO-1 for the active compounds must present lower values than the inactive compounds. This means that the active compounds are not good electron-donor molecules when compared to the inactive ones, i.e. perhaps the inactive compounds interact through a charge transfer mechanism before reaching the biological receptor, causing the loss of anti-trypanocidal activity of these compounds.

Here it is interesting to make some comments on the influence of HOMO-1 in the anti-trypanocidal activity of the quinone compounds studied in this work, since HOMO-

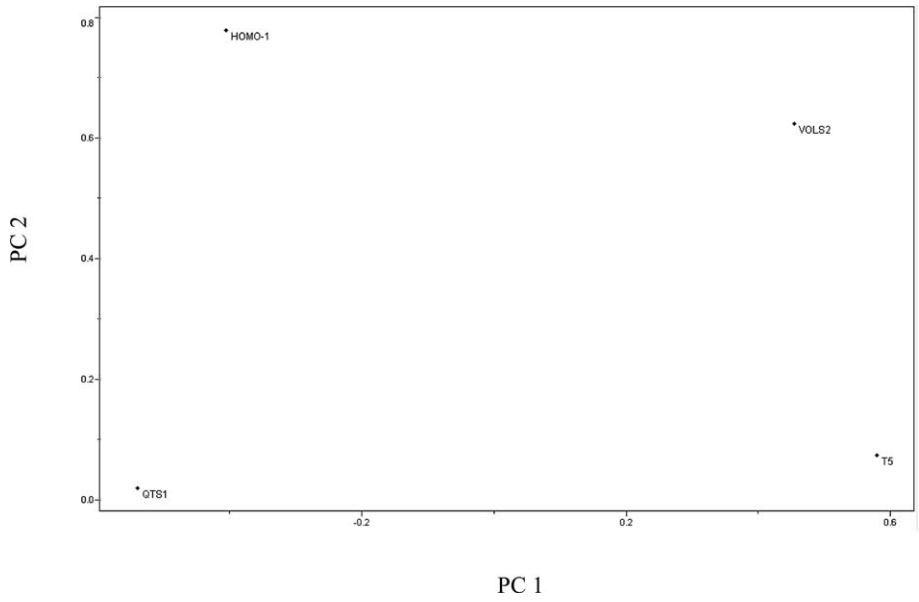

Fig. 3. Plot of the first two loading vectors (PC_1 and PC_2) of the variables responsible for the separation of the active and inactive compounds.

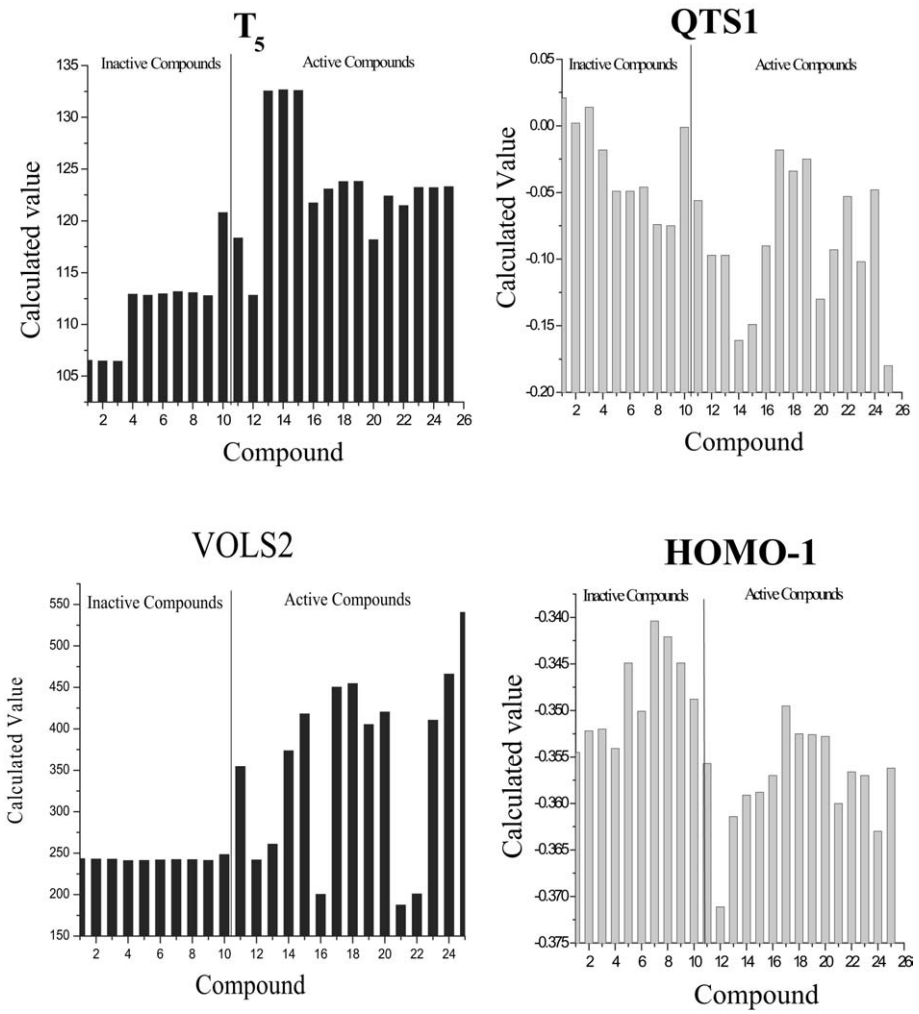

Fig. 4. Histogram with the calculated values obtained for the variables (T_5, QTS1, VOLS2 and HOMO-1) responsible for the separation between active and inactive compounds.

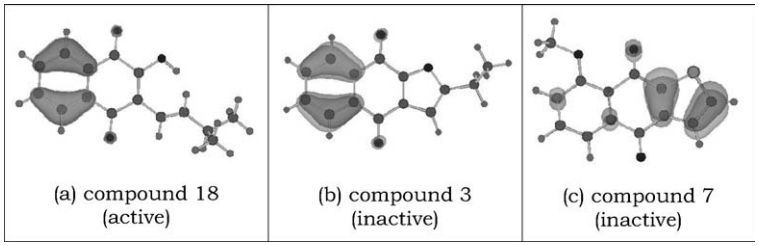

Fig. 5. Examples of the HOMO-1 contributions for three quinone compounds studied: (a) compound **18**, which represents the active compounds (we remind that all active compounds have no methoxy groups in their structures); (b) compound **3**, which represents the inactive compounds that have no methoxy groups in their structures; (c) compound **7**, which represents the inactive compounds that have methoxy groups in their structures.

1 has been also found to be an important variable in previous studies [22–25]. In order to exemplify the importance of HOMO-1 in the discrimination between the active and inactive quinone compounds studied here, we are showing in Fig. 5 where HOMO-1 has its main contributions (the three cases for the HOMO-1 behavior showed in Fig. 5 are those observed in all of the 25 quinone compounds studied). Fig. 5(a) and (b) display the HOMO-1 contributions for the active and inactive compounds, respectively, without methoxy groups in their structures; and Fig. 5(c) shows the HOMO-1 contributions for the inactive compounds that have methoxy groups in their structures. From Fig. 5 we can draw two important conclusions: (a) the active and some inactive quinone compounds have the main HOMO-1 contributions located in atoms of ring A (see Fig. 1); (b) the main HOMO-1 contributions in the inactive compounds that have a methoxy group are located in atoms of region A (see Fig. 1). So, only the inactive compounds have the HOMO-1 significantly affected by the presence or absence of methoxy groups.

Concluding, we can state that electronic and structural properties are important factors to understand the interaction between quinone compounds that present anti-trypanocidal activity and the biological receptor, where the electronic properties are related to the strength of a molecular association by electronic interaction and the structural ones are related to the positioning of the molecule during the interaction with the biological receptor. In fact, these characteristics presented by the quinone compounds studied in this work can be useful in the design of new quinone molecules with anti-trypanocidal activity.

3.2. Hierarchical cluster analysis (HCA)

HCA [26] has become, together with PCA, another important tool in multivariate data analysis. Its primary purpose is to display the data in such a way so as to emphasize its natural clustering and patterns in a two-dimensional space. The HCA results are presented in dendrograms where the vertical lines represent the compounds and the horizontal lines represent the similarity between pair of compounds, a compound and a group of compounds and groups of compounds. In HCA, the distances between samples or variables are calculated and compared through the similarity index which ranges from zero to one.

Fig. 6 shows the results obtained with our HCA analysis using the same variables selected by PCA (T_5, QTS1, VOLS2 and HOMO-1). The similarity value between the two classes (groups) of compounds was 0.0 and this means these two classes are distinct. The groups A (active compounds, **11**–**25**) and B (inactive compounds, **1**–**10**) in Fig. 6 correspond to the same groups A and B in Fig. 2 (the PCA analysis). Both PCA and HCA methods classified the 25 quinone compounds studied into two groups in exactly the same manner. Based on the classification obtained with PCA and HCA, the variables T_5, QTS1, VOLS2 and HOMO-1 are those responsible for the separation between the active and inactive anti-trypanocidal quinone compounds.

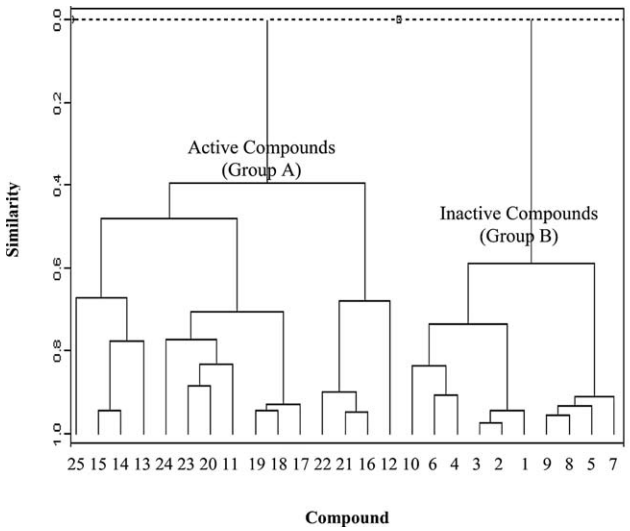

Fig. 6. Dendrogram obtained with HCA for 25 quinone compounds studied. The HCA classified the compounds into two groups: active and inactive compounds.

3.3. Stepwise discriminant analysis (SDA)

SDA is a multivariate technique that has two main goals: (1) separate objects from distinct populations and (2) allocate new objects to populations previously defined [27,28]. In this work, we considered two groups: Group A, which contains the active compounds (compounds **11**–**25** in Fig. 1) and Group B, which contains the inactive compounds (compounds **1**–**10** in Fig. 1).

The SDA is a linear discriminant method based on the Fisher's test (*F*-test) for the significance of the variables [28]. In each step one variable is selected based on its significance and after some steps the more significant variables are

extracted from the set of variables under investigation. In this work, the variables were autoscaled and after the SDA analysis the selected variables were T_5, QTS1, VOLS2 and HOMO-1 and the two discriminant functions obtained for the active and inactive quinone compounds studied are given as follows:

Group A (actives): – 0.959 + 1.335 [T_5] – 0.278 [QTS1] + 0.805 [VOLS2] – 1.438 [HOMO-1]

Group B (inactives): – 2.895 – 2.404 [T_5] + 0.325 [QTS1] – 1.454 [VOLS2] + 2.480 [HOMO-1]

It is interesting to notice that in an interaction between the active compounds and the biological receptor, the variables QTS1 and HOMO-1 represent electrostatic effects and the T_5 and VOLS2 represent steric effects. By using the quantities given in the discriminant functions above, we can obtain the classification summary showed in Table 3.

The classification error rate was 0%, resulting in a satisfactory separation of the two groups. The allocation rule derived from the SDA results, when the anti-trypanocidal activity of a new quinone compound is investigated, is: (a) initially one calculates, for the new quinone compound, the value of the more important variables obtained with the SDA methodology, i.e. T_5, QTS1, VOLS2 and HOMO-1; (b) substitute these values in the two discriminant functions obtained in this work; (c) check which discriminant function presents the higher value. The new quinone compound is active if the higher value is related to the discriminant function of Group A and vice versa.

Besides the classification matrix presented in Table 3, we also used a cross-validation test which uses the leave-one-out methodology in order to determine if the model obtained is reliable. In this procedure, one compound is omitted of the data set and the discriminant functions are built based on the remaining compounds. Afterwards, the omitted compound is classified according to the generated discriminant functions. In the next step, the omitted compound is included and a new compound is removed, and the procedure goes on until the last compound is removed. The results obtained with the cross-validation methodology are summarized in Table 4, and from these results we can see that the models obtained with PCA, HCA and SDA are reliable once the cross-validation error is equal to 0%.

Comparing the results obtained with the PCA, HCA and SDA methodologies, we can notice that the same set of variables (T_5, QTS1, VOLS2 and HOMO-1) was considered important in the three methodologies. Thus, from the results obtained with PCA, HCA and SDA we can say that T_5, QTS1, VOLS2 and HOMO-1 are key variables for explaining the anti-trypanocidal activity of the quinone compounds studied in this work and, consequently, they can be used to obtain (design) new quinone compounds with anti-trypanocidal activity.

Table 3
Classification matrix obtained with the SDA method

Classification matrix		
Classified group	True group	
	A	B
A	15	0
B	0	10
Total	15	10
Percentage	100%	100%

Table 4
Cross-validation matrix obtained with the SDA methodology

Cross-validation matrix		
Classified group	True group	
	A	B
A	15	0
B	0	10
Total	15	10
Percentage	100%	100%

3.4. Kth nearest neighbor (KNN)

The KNN method [26,29] classifies the objects based on the distance comparison among them. The multivariate Euclidian distances between every pair of samples with known class membership is calculated. The closest *K* samples are used to build the model. The optimal *K* is determined by cross-validation applied to the training set samples. The classification of a test sample is determined based on the multivariate distance of this sample with respect to the *K* samples in the training set. The classical KNN approach does not have outlier detection capability, i.e. a classification is always made whether or not the unknown object is a member of any class in the training set.

The KNN method was used in this work for the validation of the initial data set and Table 5 represents the results obtained with 1, 3, 5 and 7 nearest neighbors. For all cases, the percentage of correct information was 100%. We decided to use 7NN instead of 1NN, 3NN and 5NN because the greater the number of nearest neighbors, the better the reliability of the KNN method. The results obtained with the KNN method were similar to those obtained with PCA, HCA and SDA.

3.5. Soft independent modeling of class analogy (SIMCA)

The theoretical basis of SIMCA [26,30] is completely different from the KNN methodology. SIMCA is similar to PCA in the sense that both are based on the generation of new variables (the PCs) that are linear combinations of the original ones. However, contrary to PCA, SIMCA builds one model of PCs to each category. In this work, we performed the SIMCA analysis in order to validate the selected variables by

Table 5
Classification obtained with the KNN method

Category	Number of compounds	Compounds incorrectly classified			
		1NN	3NN	5NN	7NN
Active	15	0	0	0	0
Inactive	10	0	0	0	0
Total	25	0	0	0	0
Percentage of correct information		100%	100%	100%	100%

the PCA method, which were used for the discrimination between active and inactive quinone compounds.

The SIMCA results revealed that two PCs were needed for modeling the active compounds and three PCs were needed for the inactive compounds. Fig. 7 is a bidimensional plot of the class distances. The *F*-statistic threshold lines (defined in terms of probability and chosen here as 95%) divide the plot into four quadrants. The interclass distances and the interclass residual values are shown in Table 6. From Table 6, we can see that the SIMCA model has a good discriminating power, as the interclass distances and the interclass residual values present high values, indicating two distinct classes of compounds.

Another information obtained with the SIMCA method was the importance of the variables. For this purpose two concepts are used: modeling and discriminating power. The modeling power is the contribution of each variable to build the model, and the discriminating power is the capacity of the variables to differentiate the two classes (active and inactive compounds). Table 7 shows the modeling and discriminating power of the variables studied in this work. According to Table 7, the variable labeled as HOMO-1 has the highest value for the modeling power and QTS1 has the highest value for the discriminating power, i.e. HOMO-1 is very important for the modeling of the boxes corresponding to the two classes (active and inactive compounds) and QTS1 is the most important variable for discriminating the two classes of compounds.

Table 6
Values for the interclass distances and residuals

	Class 1		Class 2	
	Distance	Residual	Distance	Residual
Class 1	0.000	0.274	46.106	1.302
Class 2	46.106	38.974	0.000	0.781

Table 7
Modeling and discriminating power obtained with the SIMCA methodology

Variable	Modeling power	Discriminating power
T_5	0.528	652.214
QTS1	0.609	3363.775
VOLS2	0.554	1207.151
HOMO-1	0.666	851.939

3.6. Prediction study

Knowing the performance of the five pattern recognition methods (PCA, HCA, SDA, KNN and SIMCA), with the 25 quinone compounds studied in this work, we decided to apply them to a series of new quinone compounds which have similar chemical structure to the ones of our training set and whose activity against *T. cruzi* were well-known [10,11]. In fact, the opportunity of applying our methodology in a new set of quinone compounds, whose activity against *T. cruzi* is already known previously, would be a good chance to validate the models we had obtained with our training set.

Fig. 8 shows the chemical structure of the three new quinone compounds (numbered **26**–**28**). The calculated properties for the three new compounds are shown in Table 8. In order to verify if these new molecules would be active or inactive against *T. cruzi*, we applied the results obtained with the five pattern recognition methods (PCA, HCA, SDA, KNN and SIMCA) for the 25 molecules of our training set. It is interesting to emphasize that the calculated values obtained for T_5, QTS1, VOLS2 and HOMO-1, for the compounds **26**–**28**, were autoscaled in order to perform all chemometric analysis.

From the prediction results presented in Table 9, we can see that compound **26** was classified correctly as inactive and

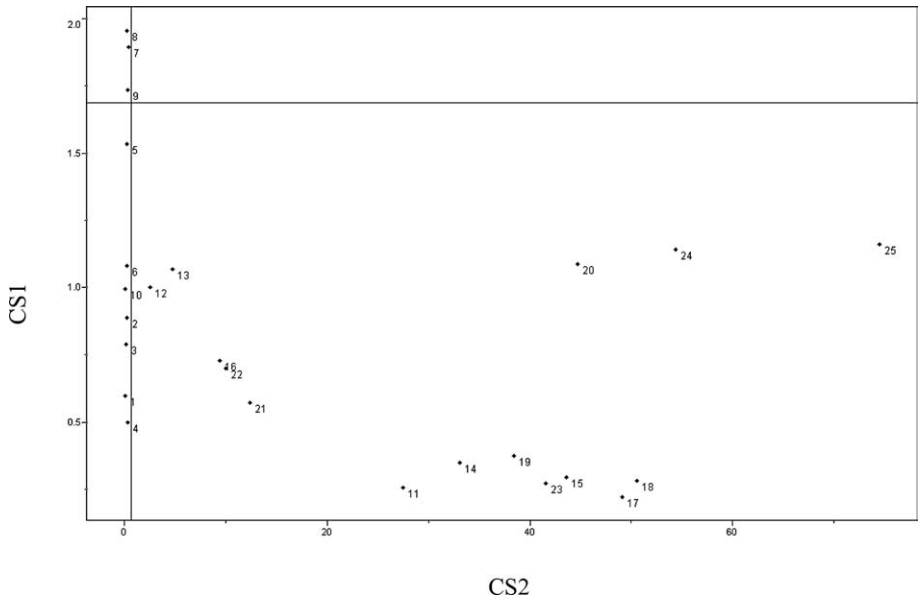

Fig. 7. Class distance plot for active (CS1) and inactive (CS2) classes of the 25 quinone compounds studied.

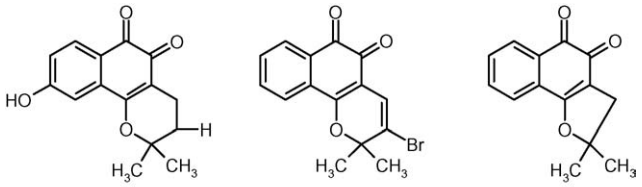

Fig. 8. Chemical structure of the three new quinone compounds (test set).

Table 8
Values obtained for the properties (variables) of the three new quinone compounds

Compound	T_5 (°)	QTS1	VOLS2 (Å^3)	HOMO-1 (eV)
26	122.93	–0.115	411.17	–0.3423
27	120.07	–0.048	465.14	–0.3622
28	128.48	–0.182	380.90	–0.3576

Table 9
The prediction results obtained with the five pattern recognition methods for the three new quinone compounds (test set): active compound (+) and inactive compound (–)

Compound	Methods				
	PCA	HCA	SDA	KNN	SIMCA
26	–	–	–	–	–
27	+	+	+	+	+
28	+	+	+	+	+

compounds **27** and **28** were classified correctly as active by using the five methods (PCA, HCA, SDA, KNN and SIMCA). Thus, we can conclude that the models obtained in this work for our training set, with the PCA, HCA, SDA, KNN and SIMCA methodologies, are reliable and can be applied to new quinone compounds whose trypanocidal activity is unknown.

4. Conclusions

In this study, theoretical calculations and chemometric methods were successfully used for modeling and predicting the behavior of 25 quinone compounds (training set) regarding their anti-trypanocidal activity. Two electronic (QTS1 and HOMO-1) and two structural properties (T_5 and VOLS2) were found as the best descriptors for the discrimination between the active and inactive quinone compounds studied.

The chemometric study performed in this work showed that four variables (T_5, QTS1, VOLS2 and HOMO-1) were responsible for the separation between the active and inactive compounds. The results obtained with PCA, HCA, SDA, KNN and SIMCA provided 100% of correct information, indicating that the models obtained in this work with all of the five methodologies (PCA, HCA, SDA, KNN and SIMCA) are reliable.

The PCA, HCA, SDA, KNN and SIMCA results, attained for the compounds of the training set, were applied to three new compounds with known anti-trypanocida activity. All methods reproduced the experimental data, i.e. the three samples studied were classified correctly, indicating the reliability of our models.

From these results, we can conclude that electronic and structural properties are important factors in the interaction between quinone compounds that present anti-trypanocidal activity and the biological receptor, where the electronic properties are related to the strength of a molecular association by electronic interaction and the structural ones are related to the positioning of the molecule during the interaction with the biological receptor.

Acknowledgements

The authors would like to thank CAPES and CNPq (Brazilian agencies) for the financial support.

References

[1] J. Cohen, R. Gürtler, Science 293 (2001) 694.
[2] F.J. Medrano, M.A. Wenck, J.C. Engel, S.P. Craig, J. Med. Chem. 46 (2003) 2548.
[3] L.S. Chemin, E. Buisine, V. Yardley, S. Kohler, M.A. Debreu, V. Landry, C. Sergheraert, S.L. Croft, R.L.K. Siegel, E.D. Charvet, J. Med. Chem. 46 (2001) 548.
[4] J.L. Howland, Biochim. Biophys. Acta 73 (1963) 665.
[5] J.L. Howland, Biochim. Biophys. Acta 77 (1963) 657.
[6] W.D. Wilson, R.L. Jones, in: Advances in Pharmacology and Chemotherapy, vol. 18, Academic Press, New York, 1981, pp. 177.
[7] H.W. Moore, J.O. Karlsson, Recents Advances in Phytochemistry, vol. 20, Plenum, New York, 1986 p. 263.
[8] M.O.F. Goulart, V. Zani, J. Tonholo, L.R. Freitas, F.C. de Abreu, A.B. Oliveira, et al., Bioorg. Med. Chem. Lett. 7 (1997) 2043.
[9] S. Subramanian, Estudo da Relação Estrutura-Atividade Biólogica do Lapachol, das Lapachonas e de alguns de seus derivados, Ph.D. Thesis, 1996, p. 253.
[10] K.C.G. de Moura, F.S. Emery, C.N. Pinto, M.C.F.R. Pinto, A.P. Dantas, K. Salomão, S.L. de Castro, A.V. Pinto, J. Braz. Chem. Soc. 12 (2001) 325.
[11] A.D. Becke, J. Chem. Phys. 98 (1993) 5648.
[12] A.D. Becke, Phys. Rev. A. 38 (1998) 3098.
[13] C. Lee, W. Yang, R.G. Parr, Phys. Rev. B 37 (1988) 785.
[14] A.A. El-Azhary, H.U. Sutter, J. Phys. Chem. 100 (1996) 15056.
[15] F. Turecek, J. Phys. Chem. 102 (1998) 4703.
[16] M.J. Frisch, G. Trucks, M. Head-Gordon, P.M.W. Gill, B.G. Johnson, M.A. Robb, et al., M. HeadGordon, C. Gonzalez, J. Pople, A. GAUSSIAN 98, Gaussian, Inc., Pittsburgh, PA, 1998.
[17] C.M. Breneman, K.B. Wiberg, J. Comput. Chem. 11 (1990) 361–373.

[18] WHIM/3D-QSAR, version 3.3, Talete srl, Milano, Italy, 1997.
[19] R. Todeschini, P. Gramatica, Perspect. Drug Discov. Des. 9 (1998) 355.
[20] G. Boon, W. Langenaeker, F. De Proft, J.P. Hans De Winter, P. Tollenaere, Geerlings, J. Phys. Chem. A 105 (2001) 8805.
[21] J.C. Pinheiro, M.M.C. Ferreira, O.A.A.S. Romero, J. Mol. Struct. 572 (2001) 35.
[22] L.L. Do E. Santo, D.S. Galvão, J. Mol. Struct. THEOCHEM 464 (1999) 273.
[23] P.M.V.B. Barone, A. Camilo Jr., D.S. Galvão, Phys. Rev. Lett. 77 (1996) 1186.
[24] R.S. Braga, P.M.V.B. Barone, D.S. Galvão, J. Mol. Struct. THEOCHEM 464 (1999) 257.
[25] G. Rothenberg, Y. Sasson, Tetrahedron 55 (1999) 561.
[26] K.R. Beebe, R.J. Pell, M.B. Seasholtz, Chemometrics: A Practical Guide, Wiley, New York, 1998 p. 348.
[27] R.A. Johnson, D.W. Wichern, Applied Multivariate Statistical Analysis, Present Hall, New Jersey, 1992 p. 594.
[28] K.V. Mardia, J.T. Kent, J.M. Bibby, Multivariate Analysis, Academic Press, New York, 1979 p. 521.
[29] B.R. Kowalski, C.F. Bender, J. Am. Chem. Soc. 94 (1972) 5632.
[30] S. Wold, Pattern Recognit. 8 (1976) 127.