

A Study of Neolignan Compounds with Biological Activity Against *Paracoccidioides brasiliensis* by Using Quantum Chemical and Chemometric Methods

Ademir J. Camargo^a, Káthia M. Honório^a, Ricardo Mercadante^b, Fábio A. Molfetta^a, Cláudio N. Alves^c and Albérico B. F. da Silva^{*,a}

^a Instituto de Química de São Carlos, Universidade de São Paulo, CP 780, 13560-970 São Carlos - SP, Brazil

^b Departamento de Química, Universidade Estadual do Oeste do Paraná, Campus de Toledo, 85903-000 Toledo - PR, Brazil

^c Departamento de Química, Centro de Ciências Exatas e Naturais, Universidade Federal do Pará, CP 11101, 66075-110 Belém - PA, Brazil

Métodos quimiométricos (estatísticos) são empregados para classificar um conjunto de compostos derivados de neolignanas com atividade biológica contra a *Paracoccidioides brasiliensis*. O método AM1 (Austin Model 1) foi utilizado para calcular um conjunto de descritores moleculares (propriedades) para os compostos em estudo. A seguir, os descritores foram analisados utilizando os seguintes métodos de reconhecimento de padrões: Análise de Componentes Principais (PCA), Análise Hierárquica de Agrupamentos (HCA) e o método de K-vizinhos mais próximos (KNN). Os métodos PCA e HCA mostraram-se bastante eficientes para classificação dos compostos estudados em dois grupos (ativos e inativos). Três descritores moleculares foram responsáveis pela separação entre os compostos ativos e inativos: energia do orbital molecular mais alto ocupado (E_{HOMO}), ordem de ligação entre os átomos C_1 - R_7 (L_{14}) e ordem de ligação entre os átomos C_5 - R_6 (L_{22}). Como as variáveis responsáveis pela separação entre compostos ativos e inativos são descritores eletrônicos, conclui-se que efeitos eletrônicos podem desempenhar um importante papel na interação entre receptor biológico e compostos derivados de neolignanas com atividade contra a *Paracoccidioides brasiliensis*.

Chemometric (statistical) methods are employed to classify a set of neolignan compounds with activity against *Paracoccidioides brasiliensis*. The AM1 (Austin Model 1) method was used to calculate a set of molecular descriptors (properties) of the neolignan compounds under study. The descriptors were further analyzed using Principal Components Analysis (PCA), Hierarchical Cluster Analysis (HCA) and K-Nearest Neighbor (KNN) chemometric methods. The PCA and HCA methods were quite efficient to classify the compounds into two groups (active and inactive) and three descriptors were found to be important in the classification: the highest occupied molecular orbital energy (E_{HOMO}), the bond order between C_1 - R_7 atoms (L_{14}) and the bond order between C_5 - R_6 atoms (L_{22}). The three variables responsible for the separation between active and inactive compounds are all electronic descriptors, therefore we can conclude that electronic effects should have an important role when one is trying to understand the activity of neolignan compounds against *Paracoccidioides brasiliensis*.

Keywords: neolignans, *Paracoccidioides brasiliensis*, AM1, PCA, HCA, KNN

Introduction

Paracoccidioides brasiliensis is a human pathogenic fungus that constitutes a major medical problem in Latin America. The microorganism is the etiological agent of paracoccidioidomycosis (PCM), a systemic disease with a

high incidence among the rural population of Latin America.¹ This disease is usually chronic and involves several organs and systems with predominance in the lungs which are considered as the primary site of the infection.² Effective treatment regimens are available to control the infectious process and most patients (60-80%) develop fibrotic sequelae that may severely hamper respiratory functions and limit the patient well-being.² Antifungal

* e-mail: alberico@iqsc.usp.br

medications are the mainstay of treatment for PCM because their mechanism of action may involve an alteration of RNA or DNA metabolism or an intracellular accumulation of peroxide that is toxic to the fungal cell.³

Some experimental and clinical investigations have indicated the relevance of humoral and/or cellular immune responses in the pathogenesis and evolution of PCM. Specific cell-mediated immune responses seem to play an important role in the resistance to *P. brasiliensis*. Patients with systemic PCM tend to show depressed cellular immune responses compared to those with localized disease. Also, the most severe forms of infection are associated with high levels of specific antibodies.⁴

Several compounds are used for inhibition of the *P. brasiliensis* and the neolignan compounds have been used for this purpose.⁵ Neolignans are organic dimers derived from oxidative coupling of allyl and propenyl phenols that occur in the *Myristicaceae*⁶⁻⁸ and other primitive plant families. The genus *Virola* presents about 35 species distributed in the neo-tropical areas occurring mainly in the Amazon rain forest.⁹⁻¹¹ The resin of several *Virola* supplies a hallucinogen powder used in rituals,¹² and the genus *Virola* is a source of lignans and neolignans with recognized bioactivity.¹³

The aim of the present work is to investigate the relationship between molecular properties and the activity against *P. brasiliensis* of synthetic neolignans and analogues by using chemometric methods.

Methodology

Quantum chemical analysis

Neolignans are molecules that present several degrees of rotation with the possibility of attaining many geometric conformations (Figure 1 shows the general structure and numbering used in the neolignan compounds studied). In the absence of crystallographic structures, it is necessary to carry out a conformational search in order to find the conformation of the lowest energy. In this work we used a careful procedure to obtain the molecular conformation associated with the lowest total energy. Initially a molecular mechanics (MM) conformational search was carried out with the Tripos 5.2 force field¹⁴ by using the software Spartan 5.0.¹⁵ Next, the final molecular conformation of each compound studied was attained by using the AM1 semi-empirical method¹⁶ employing the molecular package AMPAC 5.0.¹⁷ The molecular geometries were fully optimized by using the Precise keyword (which increases the precision of the calculations) and after the initial conformational study, the molecular properties (variables)

were calculated. The chemical structures of the 11 neolignan compounds studied in this work are present in Figure 2.

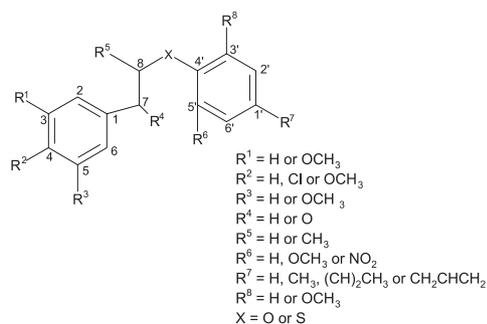


Figure 1. General structure and numbering used in the neolignan compounds studied.

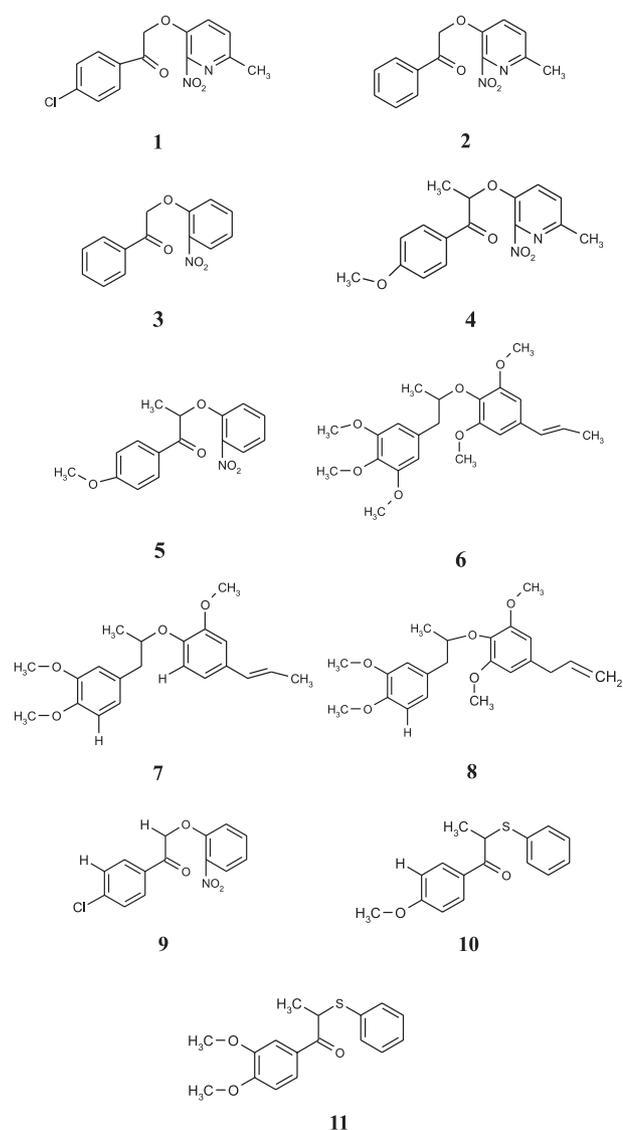


Figure 2. Active (1, 2 and 4) and inactive (3, 5, 6, 7, 8, 9, 10 and 11) neolignan compounds against *P. brasiliensis*.

Molecular properties of chemical compounds are usually correlated with biological activity. This correlation is known as structure-activity relationship (SAR)¹⁸⁻²¹ and many of such studies have been reported in the literature.²²⁻²⁴ The SAR methods have been used with success in pharmaceutical applications,²⁵ and in this work we calculated the following molecular properties to be correlated with the biological activity under study: log P: the values of this property were obtained from the hydrophobic parameters of the substituents by using the molecular package Spartan 5.0;²⁶ molecular surface area (A) and volume (V): properties evaluated with the molecular package HyperChem 5.0;²⁷ partial atomic charges (Q_n) and bond orders (L_n): derived from the NBO analysis;^{28,29} energy of the HOMO (E_{HOMO}) and LUMO (E_{LUMO}) frontier orbitals; hardness (η): obtained from the equation

$$\eta = \frac{(E_{LUMO} - E_{HOMO})}{2};$$

Mulliken electronegativity (χ): calculated from the equation

$$\chi = -\frac{(E_{HOMO} + E_{LUMO})}{2};$$

other electronic properties were calculated: total energy (E_T), heat of formation (ΔH_f); ionization potential (IP), dipole moment (μ) and polarizability (α). These values were obtained from the molecular package Ampac 5.0;¹⁷ dihedral (D_n) and bond (A_n) angles: obtained also with the Ampac 5.0 program.

Chemometric analyses

The correlation between molecular properties and biological activity was done by using the following pattern recognition methods employing the computational package Pirouette.³⁰

Principal component analysis (PCA). PCA is a multivariate statistical technique and its central idea is to reduce the dimensionality of a data set (training set) that presents a large number of interrelated variables, while retaining as much as possible the variation of the data set. This is achieved by transforming to a new set of variables, the principal components (PCs), which are uncorrelated and ordered so that the first PC retains most of the variation present in all of the original variables.¹⁹

Hierarchical Cluster Analysis (HCA). This technique examines the distances between the samples in a data set and represents this information as a two-dimensional plot called dendrogram. The HCA method is an excellent tool

for preliminary data analysis. It is informative to examine the dendrogram in conjunction with PCA as they give similar information in different forms. In HCA each point forms an only cluster initially and then the similarity matrix is analyzed. The most similar points are grouped forming one cluster and the process is repeated until all points belong to an only group.^{20,31}

K-Nearest Neighbor Analysis (KNN). The KNN method classifies a new object (compound) according to its distance to an object of the training set. The closest neighbors of the training set are found and the object will be assigned into the class that has the majority of its nearest neighbors. This method is self-validating because in the training set each sample (object) is compared with all of the objects in the set but not with itself. The best value of K can be chosen based on the results from the training set alone.³² The classical KNN approach does not have outlier detection capability, *i.e.* a classification is always made whether or not the unknown is a member of any class in the training set.

Results and Discussion

Pre-processing of the molecular descriptors

Before applying the pattern recognition methods to the 11 compounds under study each calculated property (variable or descriptor) was autoscaled. In the autoscaling method each variable is scaled to a mean of zero and a variance of unity. This method is very important because each variable is equally weighted and this provides a measure of the ability of a descriptor to discriminate classes of compounds.³² With this method we can compare all of the variables at the same level although presenting different units.

PCA and HCA analyses

After several analyses, the best separation was obtained by using the following variables: E_{HOMO} , L_{14} and L_{22} , whose calculated values are presented in Table 1. This suggests that the other variables are not significant for the classification of the compounds studied.

The PCA results show that the first component (PC_1) is responsible for 83.48% of the variance of the data. Considering the first (PC_1) and second (PC_2) components, the accumulated variance increases to 96.07%. Figure 3 shows that PC_1 is in fact responsible for the discrimination between active (1, 2 and 4) and inactive (3, 5, 6, 7, 8, 9, 10 and 11) compounds. Equation (1) presents the loading values of each variable in the PC_1 , which is responsible

Table 1. The three most important properties (descriptors) that classified the 11 compounds of the training set as active and inactive molecules

Compound	E_{HOMO} (eV)	L_{14}	L_{22}
1	-9.906	1.336	0.804
2	-10.152	1.328	0.811
3	-9.738	1.381	0.867
4	-9.504	1.323	0.811
5	-9.627	1.383	0.866
6	-8.504	1.397	1.031
7	-8.566	1.359	0.945
8	-8.398	1.392	1.041
9	-9.817	1.383	0.866
10	-8.327	1.399	0.938
11	-8.283	1.396	0.943

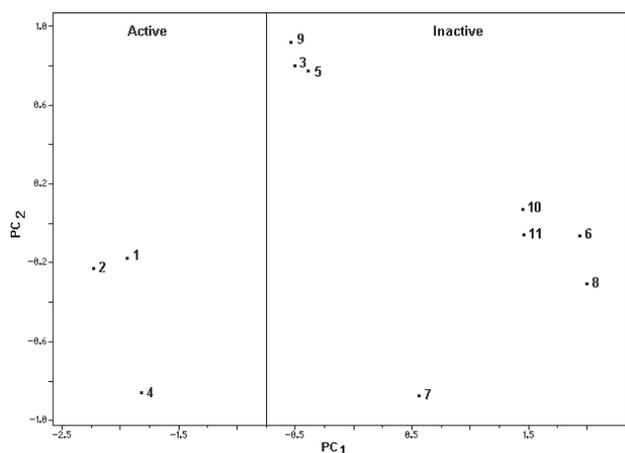


Figure 3. The separation of the training set into two groups: active and inactive compounds. Notice that the first component (PC_1) is responsible for the separation.

for the discrimination between active and inactive compounds

$$PC_1 = 0.579 [E_{\text{HOMO}}] + 0.547 [L_{14}] + 0.605 [L_{22}] \quad (1)$$

Figure 3 shows the separation of the training set of molecules into two groups: active and inactive molecules against *P. brasiliensis* when we used the variables E_{HOMO} , L_{14} and L_{22} to obtain the separation. From Figure 3 we can see that the active compounds present negative values for PC_1 while the inactive compounds present positive values for PC_1 .

From equation (1) we can see that for a compound to become active it needs to present large and negative values for the highest occupied molecular orbital energy (E_{HOMO}) along with small values for the bond order between C_1 and R_7 (L_{14}) atoms and the bond order between C_5 and R_6 (L_{22}) atoms.

It is also interesting to notice that the variables responsible for the separation between active and inactive compounds, i.e. E_{HOMO} , L_{14} and L_{22} , are all electronic variables. Therefore we can conclude that electronic effects should have an important role when one is trying to understand the activity of neolignan compounds against *P. brasiliensis*.

The energies of the frontier orbitals are important properties in several chemical and pharmacological processes. The reason for this is the fact that these properties give information on the electron donating and electron accepting character of a compound and, consequently, on the formation of a charge transfer complex (CTC). The energy of the highest occupied molecular orbital (E_{HOMO}) measures the electron donating character of a compound and the energy of the lowest unoccupied molecular orbital (E_{LUMO}) measures its electron accepting character.³³⁻³⁸ From these definitions, we have that: (a) the greater the E_{HOMO} , the greater the electron donating capability; (b) the smaller the E_{LUMO} , the smaller the resistance to accept electrons.³⁹ For the active compounds studied in this work we can say that their E_{HOMO} must present negative values whereas the inactive compounds must present positive values. This means that the inactive compounds are more efficient electron donor compounds than the active ones, i.e. the inactive compounds may interact through charge transfer mechanism with some compounds before reaching the biological receptor.

Regarding the bond order descriptor we can define it as half of the difference among electrons in bonding and anti-bonding molecular orbitals. The greater the bond order, the greater the dissociation energy and the smaller the bond length. Thus, for the active compounds studied we can conclude that some groups on the C_1 and C_5 positions are required so that the electronic density between C_1 - R_7 and C_5 - R_6 presents a low value and, consequently, a low bond order. So, the smaller the L_{14} (C_1 - R_7) and L_{22} (C_5 - R_6) bond orders, the greater the possibility of interaction between the compounds studied and the biological receptor. Figure 4 illustrates the clear separation between active and inactive compounds by using the bond orders obtained in this work (L_{14} and L_{22}), confirming the importance of these descriptors in the discrimination of neolignan compounds with biological activity against *P. brasiliensis*.

The HCA results are abridged in the dendrogram showed in Figure 5. From this dendrogram we can notice that the similarity observed between the group of the active (A) and inactive (B) molecules is small and this means that these two classes are distinct.

The HCA and PCA methods are complementary and for the 11 neolignan compounds studied in this work the

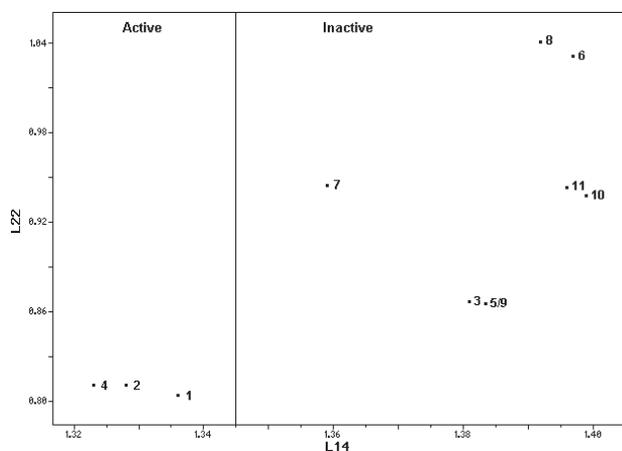


Figure 4. Plot of L_{14} versus L_{22} which illustrates the discrimination between active and inactive neolignan compounds.

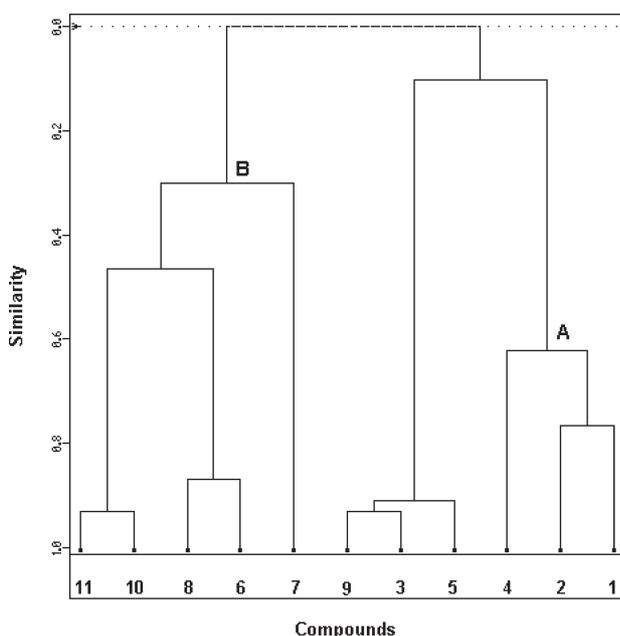


Figure 5. Dendrogram obtained with the HCA method. The training set was classified into two groups: A (active compounds) and B (inactive compounds).

HCA and PCA results were similar, i.e. HCA and PCA classified the 11 neolignan compounds exactly in the same way.

KNN analysis

The KNN method was used for the validation of the initial data set and Table 2 presents the results obtained with one (1NN), three (3NN) and five (5NN) nearest neighbors. For the case of 1NN, the percentage of correct information was 100% while for 3NN and 5NN the percentage decreased considerably (81.8 and 72.7%,

respectively). Thus, we can see that the use of 1NN leads to a higher percentage of correct information.

Table 2. Classification obtained with the KNN method

Category	Number of compounds	Compounds incorrectly classified		
		1NN	3NN	5NN
Active	3	0	2	3
Inactive	8	0	0	0
Total	11	0	2	3
Percentage of correct information		100	81.8	72.7

The KNN results were also similar to those obtained with the PCA and HCA methods. The outcomes obtained with the three classification methods (PCA, HCA and KNN) were quite interesting as we had 100% of success in classifying a data set using these three methods.

Conclusions

The application of three pattern recognition methods (PCA, HCA and KNN) on neolignan compounds with activity against *P. brasiliensis* showed that the compounds studied in this work can be correctly classified into two groups: active and inactive molecules. The PCA results showed that the variables E_{HOMO} , L_{14} and L_{22} are responsible for the separation between the active and inactive compounds. The HCA and KNN results were similar to those obtained with PCA, i.e. both methods classified the neolignan compounds exactly in same way as PCA. From the results obtained with the three chemometric (statistical) methods (PCA, HCA and KNN), we can see that the variables responsible for the separation between active and inactive compounds, i.e. E_{HOMO} , L_{14} and L_{22} , are all electronic variables. Therefore we can conclude that electronic effects should have an important role when one is trying to understand the activity of neolignan compounds against *P. brasiliensis*.

Acknowledgements

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

References

- Fonseca, C.A.; Jesuino, R.S.A.; Felipe, M.S.S.; Cunha, S.A.; Brito, W.A.; Soares, C.M.A.; *Microbes Infect.* **2001**, *3*, 535.
- Cock, A.A.; Cano, L.E.; Vélez, D.; Aristizábal, B.H.; Trujillo, J.; Restrepo, A.; *Rev. Inst. Med. Trop. S. Paulo* **2000**, *42*, 59.

3. <http://www.emedicine.com/derm/topic863.htm>, accessed in August 2002.
4. Cano, L.E.; Singer-Vermees, L.M.; Costa, T.A.; Mengel, J.O.; Xidieh, C.F.; Arruda, C.; André, D.C.; Vaz, C.A.C.; Burger, E.; Calich, V.L.G.; *Infect. Immun.* **2000**, *68*, 352.
5. Ferri, P.H.; Fonseca, C.A.; Pereira, M.; Soares, C.M.A.; Ferreira, H.D.; Nalini, A.D.G.; Oliveira, G.R.; Fernandes, E.S.; Gaspar, T.P.; *Abstracts of the XVII Congresso Brasileiro de Microbiologia*, Santos, Brazil, 1995.
6. Yoshida, M.; Gottlieb, O.R.; *Abstracts of the Brazilian Symposium on the Chemistry of Lignins and Other Wood Components*, São Carlos, Brazil, 1989.
7. Gottlieb, O.R.; *Phytochemistry* **1942**, *11*, 1537.
8. Gottlieb, O.R.; Yoshida, M.; *Quim. Nova* **1984**, *7*, 250.
9. Gottlieb, O.R.; *Fortschr. Chem. Org. Naturst.* **1977**, *35*, 1.
10. Whiting, D.A.; *Nat. Prod. Rep.* **1990**, *7*, 349.
11. Rodrigues, W. A.; *Acta Amazonica* **1980**, *10*, 1.
12. Davis, E.W.; Yost, J.A.; *J. Ethnopharmacol.* **1983**, *9*, 273.
13. Macrae, W.D.; Towers, G.H.N.; *Phytochemistry* **1984**, *23*, 1207.
14. Clark, M.; Cramer, R.D.; Opdenisch, N.V.; *J. Comput. Chem.* **1989**, *10*, 982.
15. Hehre, W.J.; Huang, W.W.; Klunzinger, P.E.; Deppmeier, B.J.; Driessen, A.J.; *Spartan 5.0; Program for molecular mechanics and quantum chemical calculations*; University of California, USA, 1997.
16. Dewar, M.J.; Zoebisch, E.G.; Healy, E.F.; Stewart, J.J.P.; *J. Am. Chem. Soc.* **1985**, *107*, 3902.
17. Dewar, M.J.; *AMPAC 5.0; Program for semiempirical calculations*; University of Texas, USA, 1994.
18. Karcher, W.; Devillers, J.; *Chemical and Environmental Science*, Kluwer Academic Publishers: Dordrecht, 1990.
19. Jolliffe, I.T.; *Principal Component Analysis*, Springer-Verlag: New York, 1986.
20. Kowalski, B.R.; Bender, C.F.; *J. Am. Chem. Soc.* **1972**, *94*, 5632.
21. Mardia, K.V.; Kent, J.T.; Bibby, J.M.; *Multivariate Analysis*, Academic Press: New York, 1979.
22. Alves, C.N.; Barroso, L.P.; Santos, L.S.; Jardim, I.N.; *J. Braz. Chem. Soc.* **1998**, *9*, 577.
23. Alves, C.N.; de Macedo, L.G.M.; Honório, K.M.; Camargo, A.J.; Santos, L.S.; Jardim, I.N.; Barata, L.E.S.; da Silva, A.B.F.; *J. Braz. Chem. Soc.* **2002**, *13*, 300.
24. Camargo, A.J.; Mercadante, R.; Honório, K.M.; Alves, C.N.; da Silva, A.B.F.; *J. Mol. Struct. (Theochem)* **2002**, *583*, 105.
25. Kubinyi, H.; *QSAR: Hansch Analysis and Related Approaches*, VCH: New York, 1993.
26. Nys, G.G.; Rekker, R.F.; *Eur. J. Med. Chem.* **1974**, *9*, 361.
27. Ostlund, N. S.; *HyperChem 4.5; Program for molecular visualization and simulation*; University of Waterloo, Canada, 1995.
28. Carpenter, J.E.; Weinhold, F.; *THEOCHEM* **1988**, *169*, 41.
29. Carpenter, J.E.; *Ph.D. Thesis*, University of Wisconsin-Madison, USA, 1987.
30. Rohrbach, B.; *Pirouette; Program for Multivariate Data Analysis*; Infometrix Inc., Seattle, WA, 1996.
31. Johnson, R.A.; Wichern, D.W.; *Applied Multivariate Statistical Analysis*, Prentice-Hall: New Jersey, 1992.
32. Lindon, J.C.; Holmes, E.; Nicholson, J.K.; *Prog. Nucl. Magn. Reson. Spectrosc.* **2001**, *39*, 1.
33. Korolkovas, A.; *Fundamentos da Farmacologia Molecular*, Guanabara Dois S.A.: Rio de Janeiro, 1982.
34. Clare, B. W.; *Theor. Chim. Acta* **1994**, *87*, 415.
35. Da Silva, A. B. F.; *M. Sc. Thesis*, Universidade de São Paulo, Brazil, 1985.
36. Ariens, E. J.; *Drug Design*; Academic Press: New York, 1971.
37. Kier, L. B.; *Molecular Orbital Studies in Chemical Pharmacology*; Springer: Berlin, 1970.
38. Burger, A.; *Medicinal Chemistry*; Wiley-Interscience: New York, 1970.
39. Honorio, K.M.; da Silva, A.B.F.; *Int. J. Quantum Chem.* **2003**, *95*, 126.

Received: November 8, 2002

Published on the web: September 29, 2003

FAPESP helped in meeting the publication costs of this article.