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-vizinhas mais próximas (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 ($E_{\text{HOMO}}$), ordem de ligação entre os átomos $C_{1'}-R_7$ (L14) e ordem de ligação entre os átomos $C_{5'}-R_6$ (L22). 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 ($E_{\text{HOMO}}$), the bond order between $C_{1'}-R_7$ atoms (L14) and the bond order between $C_{5'}-R_6$ atoms (L22). 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 paracoccidiomycosis (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
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.\textsuperscript{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 \textit{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.\textsuperscript{4}

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

The aim of the present work is to investigate the relationship between molecular properties and the activity against \textit{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 field\textsuperscript{14} by using the software Spartan 5.0.\textsuperscript{15} Next, the final molecular conformation of each compound studied was attained by using the AM1 semi-empirical method\textsuperscript{16} employing the molecular package AMPAC 5.0.\textsuperscript{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.](image)

![Figure 2. Active (1, 2 and 4) and inactive (3, 5, 6, 7, 8, 9, 10 and 11) neolignan compounds against \textit{P. brasiliensis}.](image)
Molecular properties of chemical compounds are usually correlated with biological activity. This correlation is known as structure-activity relationship (SAR) and many of such studies have been reported in the literature. The SAR methods have been used with success in pharmaceutical applications, 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; molecular surface area (A) and volume (V); properties evaluated with the molecular package HyperChem 5.0; partial atomic charges (Qn) and bond orders (Ln); derived from the NBO analysis; energy of the HOMO (E_{HOMO}) and LUMO (E_{LUMO}) frontier orbitals; hardness (\eta): obtained from the equation
\[ \eta = \frac{(E_{LUMO} - E_{HOMO})}{2}; \]
Mulliken electronegativity (\chi): calculated from the equation
\[ \chi = \frac{(-E_{HOMO} + E_{LUMO})}{2}; \]
other electronic properties were calculated: total energy (E_{T}); heat of formation (\Delta H_f); ionization potential (IP), dipole moment (\mu) and polarizability (\alpha). These values were obtained from the molecular package Ampac 5.0; 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.

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.

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.

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. 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. 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 (PC1) is responsible for 83.48% of the variance of the data. Considering the first (PC1) and second (PC2) components, the accumulated variance increases to 96.07%. Figure 3 shows that PC1 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 PC1, 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

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

Figure 3 shows the separation of the training set into two groups: active and inactive compounds. Notice that the first component (PC1) is responsible for the separation.

It is also interesting to notice that the variables responsible for the separation between active and inactive compounds, i.e. \( E_{\text{HOMO}} \), \( L_{14} \), 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 \( (E_{\text{HOMO}}) \) measures the electron donating character of a compound and the energy of the lowest unoccupied molecular orbital \( (E_{\text{LUMO}}) \) measures its electron accepting character. From these definitions, we have that: (a) the greater the \( E_{\text{HOMO}} \) the greater the electron donating capability; (b) the smaller the \( E_{\text{LUMO}} \) the smaller the resistance to accept electrons. From the active compounds studied in this work we can say that their \( E_{\text{HOMO}} \) 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
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

<table>
<thead>
<tr>
<th>Category</th>
<th>Number of compounds</th>
<th>Compounds incorrectly classified</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1NN</td>
<td>3NN</td>
</tr>
<tr>
<td>Active</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>Inactive</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>11</td>
<td>0</td>
</tr>
<tr>
<td>Percentage of correct information</td>
<td>100</td>
<td>81.8</td>
</tr>
</tbody>
</table>

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 $E_{\text{HOMO}}$, $L_{14}$ 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 the 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. $E_{\text{HOMO}}$, $L_{14}$ 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**

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.
17. Dewar, M.J.; AMPAC 5.0; Program for semiempirical calculations; University of Texas, USA, 1994.
27. Oslund, N. S.; HyperChem 4.5; Program for molecular visualization and simulation; University of Waterloo, Canada, 1995.

Received: November 8, 2002
Published on the web: September 29, 2003

FAPESP helped in meeting the publication costs of this article.