Structure-Activity Relationship of Compounds which are Anti-Schistosomiasis Active

Cláudio Nahum Alves[a]*, Lúcia Pereira Barroso[b], Lourivaldo S. Santos[a], and Iselino Nogueira Jardim[a]

[a]*Departamento de Química, Centro de Ciências Exatas e Naturais, Universidade Federal do Pará, C.P. 11101, 66075-110 Belém - PA, Brazil*

[b]*Departamento de Estatística, Instituto de Matemática e Estatística de São Paulo, Universidade de São Paulo, C.P. 66281, 05315-970 São Paulo - SP, Brasil*

O método do orbital molecular AM1 foi empregado para calcular um conjunto de descritores moleculares para vinte neolignanas sintéticas com atividade anti-esquistossomose. O método de reconhecimento de padrão (análise de componentes principais ACP, análise de conglomerados AC e análise de discriminante) foi utilizado para obter a relação entre a estrutura molecular e a atividade biológica. O conjunto de moléculas foi classificado em dois grupos de acordo com seus graus de atividade biológica. Estes resultados permitem que, projete-se racionalmente novos compostos, potenciais candidatos à síntese e à avaliação biológica.

The molecular orbital method AM1 was employed to calculate a set of molecular descriptors for twenty synthetic neolignan derivatives with anti-schistosomiasis activity. The pattern recognition method (principal component analysis PCA and cluster analysis CA and discriminant analysis methods) have been employed to obtain the relationship between the molecular descriptors and biological activity. The set of molecules was classified into two groups according to their degree of biological activity. These results allow us to rationally project new compounds, potential candidates for synthesis and biological evaluation.

Keywords: *AM1, neolignans, anti-schistosomiasis, molecular descriptors, pattern recognition method*

Introduction

Neolignans are dimers formed from oxidative coupling of allyl and propenyl phenols that occur in the Myristicaceae and other primitive plant families. The *Virola* is the most representative Myristicaceae occurring throughout the Americas[1,2]. In 1970, initial studies of leaves of *Virola surinamensis*[3] showed high efficacy in the blockage tests of penetration of cercariae of *Schistosoma mansoni* in mice. The active substances responsible for protection were isolated and identified as the natural neolignans virolin and surinamensin (Fig. 1).

In order to determine the biological activity of neolignans, Barata *et al.*[4] and Santos[5] synthesized twenty ana-logues of the compounds in (Fig. 1), which were submitted to biological tests against fungi, bacteria, leishmaniasis, schistosomiasis, cancer and PAF (platelet activating factor)[6]. Of the twenty compounds tested in vitro, seven have been classified as active and thirteen have been classified as inactive (Fig. 2) against schistosomiasis.

The structure-activity relationship (SAR) is still a basic method in molecular modeling. As the importance of three-dimensional microscopic interaction and binding between a substrate and a receptor increases, the importance of the quantum chemical parameters increases in the SAR analysis. The quantum chemical quantities of molecules and of interacting molecular systems can in principle express all electronic properties relating to molecular interactions. The

*Corresponding author (e-mail: claudio@ufpa.br)
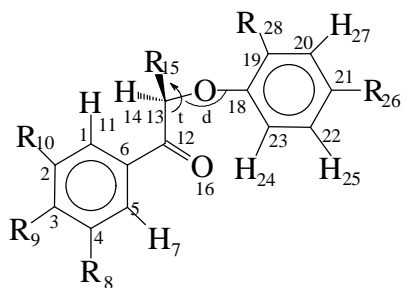
a) Virolin; R8=H, R9=R10=R28=-OCH$_3$,
   R15=-CH$_3$, R26=-CH=CH-CH$_3$ (trans)

b) Surinamensin;R8=R9=R10=R28=-OCH$_3$,
   R15=-CH$_3$, R26=-CH=CH-CH$_3$ (trans)

Figure 1. The natural neolignans, virolin (a) surinamensin (b), were isolated from leaves of *Virola surinamensis* and they showed biological activity against *Schistosoma mansoni*. The twenty neolignan analogues were synthesized modifying substituents, Rs, and position 17; t$_1$ and t$_2$ are the dihedral angles; and d the bond angle.

SAR with quantum chemical descriptors have become important in quantitative analysis of three-dimensional molecular interactions[7].

The present work employs the semi-empirical AM1 method to calculate selected molecular descriptors of twenty neolignans (Fig. 2) synthetized by Santos[5]. The pattern recognition method has been employed to obtain the quantitative relationship between the molecular descriptors and biological activity.

Methodology

Quantum chemical descriptors

The general methodology employed in the present study consists of calculation of structural descriptors by a semi-empirical quantum mechanical method followed by statistical analysis of these descriptors (Table 1). Many quantum chemical descriptors have been utilized in SAR studies[8-14]. In this work we have calculated the following descriptors:

ε$_{HOMO}$ - absolute eigenvalue of the highest occupied molecular orbital in eV;

ε$_{LUMO}$ - absolute eigenvalue of the lowest unoccupied molecular orbital in eV;

χ - Mulliken's electronegativity in a.u.;

μ - dipole moment in a.u.;

POL - molecular polarizability in a.u.;

ΔH$_f$ - heat of formation in kcal mol^{-1};

Q$_N$ - net atomic charge on atom N;

d - bond angle;

t - torsion angle;

All of the descriptors were obtained using the Hamiltonian AM1 of the version 6.0 of the MOPAC Program[15] running on an IBM RISC 6000. The statistical analysis was done with the MINITAB10.1 Program[16]. The geometries

were obtained with the molecular mechanics method MM2[17] and optimized with the semi-empirical AM1 method.

Results and Discussion

Principal component analysis (PCA)

The central idea of PCA is to reduce the dimensionality of the data set, explaining the variance-covariance structure. This is achieved by linear transformation of the original data set of variables into a smaller number of uncorrelated significant principal components (PCs). Geometrically, this transformation represents the rotation of the original coordinate system. The direction of the maximum residual variance is given by the first principal component axis; the second principal component, orthogonal to the first one, has the second maximum variance and so on. In this way, projections conserving maximum amounts of statistical information can be plotted using micro computers in order to show us a more detailed study of the properties structure[18].

A number of scores plots were examined and the most informative ones are presented in (Fig. 3) by first principal component against the second component. This projection conserves 51% of the total variance of the original data and can be expected to provide a reasonably accurate representation of the higher order space[19,20].

(Fig. 3) shows the separation of the compounds into two groups, A and B. Group A consists of five experimentally active compounds (6, 7, 10, 11 and 19). Group B consists of thirteen inactive compounds (3, 4, 5, 8, 9, 12, 13, 14, 15, 16, 17, 18 and 20) as well as the two remaining active compounds (1 and 2).

Hierarchical clustering methods

The basic aim of cluster analysis is to find 'natural grouping', if any, of a set of individuals (or objects, or points, or units, or whatever). This set of individuals may form a complete population or be a sample from some larger population. More formally, Cluster analysis aims to allocate a set of individuals to a set of mutually exclusive, exhaustive, groups such that individuals within a group are similar to one another while individuals in different groups are dissimilar. In the complete-link method, the 'distance' between two groups is defined to be the dissimilarity between the most remote pair of individuals[18].

The hierarchial dendrogram, Fig. 4, can also be used to provide information on chemical behavior and verify the results obtained by PCA. The horizontal lines represent the compounds, and the vertical lines the similarity values between pairs of compounds, a compound and a group of compounds, and between groups of compounds.

In comparison of the hierarchical cluster analysis results with those obtained from PCA, a good correlation can
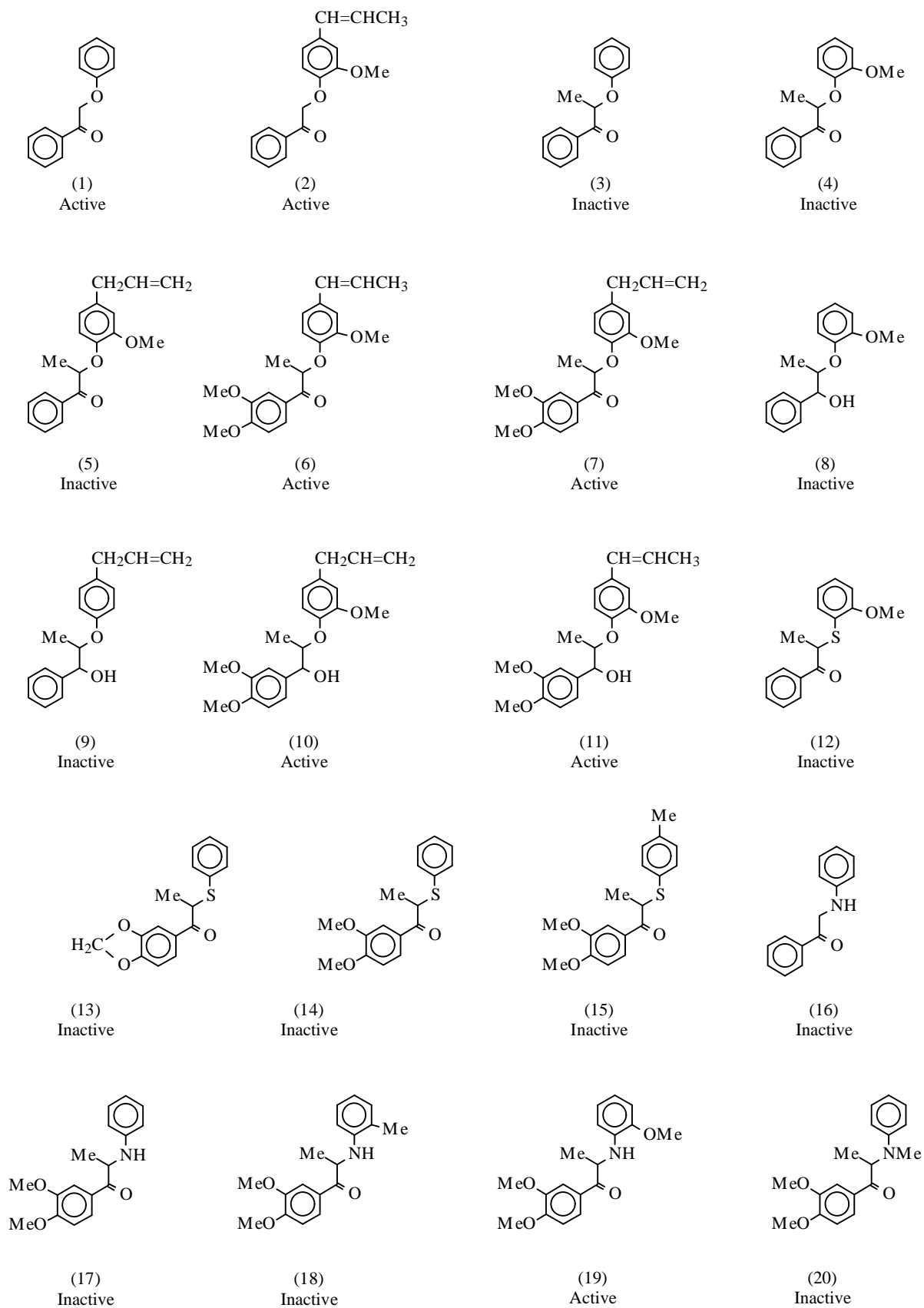
Figure 2. Synthetic neolignans, active and inative against schistosomiasis.

Table 1. The values of the 14 descriptors of the twenty compouds.

| Compound	ε_{HOMO}	ε_{LUMO}	χ	μ	d	t	ΔH_f	POL	Q2	Q3	Q12	Q13	Q19	Q21
1	9.0670	-0.5076	4.2797	3.8200	116.00	-176.40	-12.890	130.42	-0.1418	-0.0987	0.2641	-0.0703	-0.1505	-0.2072
2	8.5984	-0.4090	4.0998	3.0625	115.44	167.97	-38.618	178.88	-0.1421	-0.1004	0.2625	-0.0719	0.0652	-0.0355
3	9.1903	-0.4183	4.3909	1.7000	115.86	89.540	-18.020	135.85	-0.1491	-0.0963	0.2492	-0.0101	-0.1483	-0.1542
4	9.8142	-0.3479	4.2311	2.4200	115.10	89.920	-52.890	152.12	-0.1508	-0.0969	0.2481	-0.0127	0.0655	-0.1155
5	9.6514	-0.3377	4.1562	2.5700	116.08	87.200	-40.340	179.86	-0.1516	0.0968	0.2480	-0.0132	0.0619	0.0590
6	9.6504	-0.4520	4.0990	2.8000	115.54	91.310	-119.00	221.21	-0.0100	0.1052	0.2528	-0.0110	0.0695	-0.0277
7	8.8175	-0.4410	4.1895	3.1000	115.74	93.040	-113.89	212.99	-0.0080	0.1042	0.2554	-0.0132	0.0681	-0.0567
8	8.8052	0.2672	4.5336	2.8800	116.17	100.80	-68.680	152.94	-0.1374	-0.1251	0.0544	0.0030	0.0646	-0.1178
9	7.7336	0.4095	0.0000	2.5192	116.67	179.27	-3.283	159.05	-0.1318	-0.1244	0.0791	0.0038	-0.1929	-0.1769
10	8.3562	0.2024	4.2812	3.5870	116.38	97.430	-129.66	215.86	0.0486	0.0602	0.0569	0.0033	0.0644	-0.0563
11	8.4466	0.1741	4.3101	3.8300	116.34	93.140	-128.68	213.91	-0.1704	0.0476	0.0520	0.0037	0.0601	-0.0600
12	8.2711	-0.4710	3.8995	1.8750	108.15	72.850	-17.480	146.92	-0.1470	-0.0982	0.2764	-0.2825	-0.1193	-0.1340
13	8.3221	-0.6514	3.8353	1.7700	107.61	68.570	-40.560	164.96	-0.0096	0.0439	0.2781	-0.2837	-0.1203	-0.1344
14	8.3158	-0.4676	3.9212	3.0860	108.49	64.700	-55.240	179.13	-0.0054	0.1025	0.2803	-0.2806	-0.1223	-0.1348
15	8.2131	-0.4530	3.8785	3.0300	108.33	63.190	-62.850	189.82	-0.0064	0.1026	0.2804	-0.2800	-0.1181	-0.0732
16	8.5134	-0.5638	3.9731	3.0700	118.03	145.78	-20.920	134.17	-0.1429	-0.0978	0.2590	-0.0987	-0.1855	-0.1682
17	8.5873	-0.3250	4.1310	4.2100	118.81	81.860	-51.320	172.33	-0.0155	0.1033	0.2755	0.0431	-0.1875	-0.1657
18	8.4856	-0.3064	4.0898	4.5200	119.53	77.110	-58.110	180.47	-0.0131	0.1025	0.2756	-0.0399	-0.1241	-0.1638
19	8.4065	-0.3580	4.0210	4.4970	118.30	33.740	-88.400	181.73	-0.0021	0.0954	0.2470	-0.0461	0.0215	-0.1301
20	8.7855	-0.2511	4.2670	4.0200	116.39	54.870	-41.540	175.80	-0.0189	0.1033	0.2788	0.0442	-0.1416	-0.1450
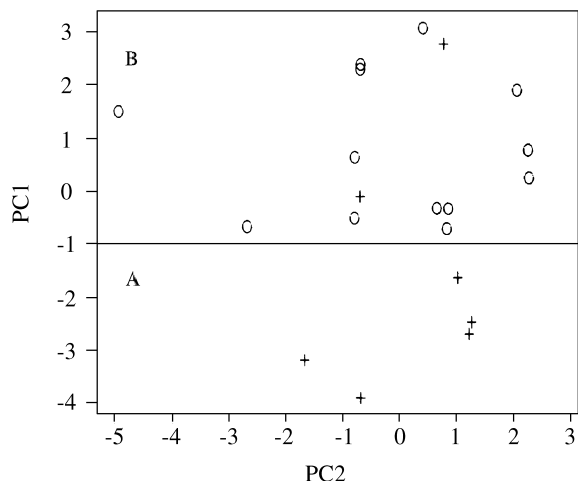
Figure 3. PC$_1$ *vs.* PC$_2$ scores plot obtained from calculated molecular descriptors; active (+) and inactive (°).
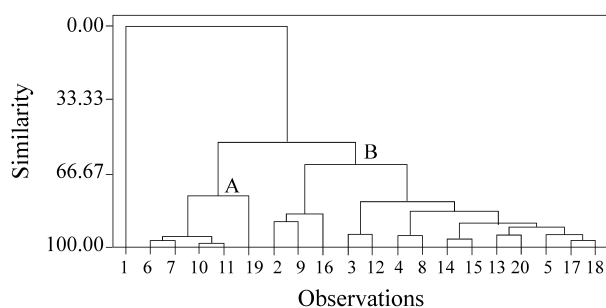

Figure 4. Dendrogram obtained with the complete-link method.

be seen coating in the dendrogram clusters and in the graph of the values (scores) of the first two principal components. We can see two groups, A and B. Group A consists of five active compounds (6, 7, 10, 11 and 19). Group B consists of thirteen inactive compounds (9, 16, 3, 12, 4, 8, 14, 15, 13, 20, 5, 17 and 18) and one active compound (2). The compound (1) was not found in either of the groups.

Discriminant analysis

Discriminant Analysis is a multivariate technique that has two principal objectives: (1) separating objects from distinct populations; (2) allocating new objects to populations previously defined. In this paper, we consider two populations: Population A consists of the active compounds (1, 2, 6, 7, 10, 11 and 19); Population B consists of the inactive compounds (3, 4, 5, 8, 9, 12, 13, 14, 15, 16, 17, 18 and 20).

Discriminant Analysis for this data set presented the following linear discriminant functions:

Population A: -19597 + 975 Homo + 10 ΔH$_f$ - 422 μ + 244 d – 2 t + 1 χ + 23 POL + 1441 Lumo – 693 Q$_2$ + 1352 Q$_3$ + 1443 Q$_{12}$ – 8358 Q$_{13}$ + 1563 Q$_{19}$ – 2976 Q$_{21}$

Table 2. Classification matrix.

Put into pop.	True Population	
	A	B
A	7	0
B	0	13
Total	7	13
Percent.	100%	100%

Population B: - 19546 + 958 Homo + 10 ΔH$_f$ - 440 μ + 246 d – 2t + 8 χ + 21 POL + 1567 Lumo – 750 Q$_2$ + 1648 Q$_3$ + 1747 Q$_{12}$ – 8421 Q$_{13}$ + 1505 Q$_{19}$ – 2709 Q$_{21}$

Using these quantities, the classification matrix is given in Table 2. The error rate 0% (classification). The separation of the two groups is quite good. The allocation rule is: calculate the two linear discriminant functions applied to the new observation; this observation is allocated to the population with the largest value.

Conclusion

The use of quantum mechanical descriptors represents a good option in SAR studies based on a semi-empirical method since they permit the prediction of a great number of molecular properties.

In our study, the PCA shows that the compounds can be classified into two groups (A and B) according to their degree of biological activity, and cluster analyses is used to verify the results obtained by PCA. The results of the hierarchical clustering method are similar to the PCA. Discriminant analysis shows the two groups have a good separation and gives a rule to classify a new compound.

The results allow us to rationally project new compounds, potential candidates for synthesis and biological evaluation.

References

1. Gottlieb, O.R.; Fortschr. *Chem. Org. Naturst.* **1977**, *1*, 35.
2. Whiting, D.A. *Nat. Prod. Rep*. **1990**, *349*, 7.
3. Barata, L.E.S. *Isolamento e Sínteses de Neolignanas de Virola surinamensis*, Ph.D. Thesis, Universidade Estadual de Campinas, Campinas-SP, Brazil, 1976.
4. Barata, L.E.S.; Baker, P.M.; Gottlieb, O.R.; Ruveda, E.A. *Phytochemistry* **1978**, *783*, 17.
5. Santos, L.S. *Síntese e Atividade Biológica de Neolignanas 8.0.4' Derivados e Compostos Correlatos*, Ph.D. Thesis, Universidade Estadual de Campinas, Campinas-SP, Brazil, 1991.

6. Barata, L.E.S.; Santos, L.S.; Fernandes, A.M.A.P.; Ferri, P.H.; Paulo, M.Q.; Neal, R.; Joudan, M.C. In *Natural and Synthetic Bioactive Neolignans*, 2nd Brazilian-German Symposium on Natural Products, Hannover, 1991, 18.

7. Kikuchi, O. *Quant. Struct.-Act. Relat.* 1987, *179*, 6.

8. Araujo, M.T.; Carneiro, J.W.M. *J. Mol. Struct. (THEOCHEM)* **1995**, *255*, (335).

9. Nguyen-Cong, V.; Rode, B.M. *J. Chem. Imf. Comput. Sci.* **1996**, *14*, 36.

10. Silva, A.B.F.; Trsic, M. *J. Mol. Struct.(THEOCHEM)* **1995**, *274*, 356.

11. Gombar, V.K.; Enslein, K. *Quant. Struct.-Act. Relat.* **1990**, *321*, 9.

12. Hemken, H.G.; Lehmann, P.A. *Quant. Struct.-Act. Relat.* **1992**, *332*, 11.

13. Cocchi, M.C.; Menziani, M.; Fanelli, F.; Benetti, P.G. *J. Mol. Stuct. (THEOCHEM)* **1995**, *79* 331.

14. Costa, M.C.A.; Barata, L.E.S.; Takahata, Yuiji *J. Mol Struct.* (THEOCHEM) **1995**, *185* 340.

15. Stewart, J.J.P. *J. Comput.-Aided Mol. Design*. **1990**, *1* 4.

16. Ryan, B.F.; Joiner, B.L.; Ryan Jr, T.A. Minitab Handbook, 2nd. Ed., Boston: PWS-KENT, 1985.

17. Allinger, N.L. *J. Am. Chem. Soc.* **1977**, *8127*, 99.

18. Chatterjee, S.; Price, B. In *Regression Analysis by Example*, John Wiley & Sons, New York, 1977.

19. Geyer, A.L.M.; Moreira, J.C.; Faigle, J.F.; Bruns, R.E.; Curtius, A.J. *Hydrobiologia* **1990**, *143*, 19.

20. Kowalski, B.R.; Bender, C.F. *J. Am. Chem. Soc.* **1972**, *5632*, 9.