A structure–activity relationship (SAR) study of synthetic neolignans and related compounds with biological activity against Escherichia coli

ARTICLE *in* JOURNAL OF MOLECULAR STRUCTURE THEOCHEM · APRIL 2002

Impact Factor: 1.37 · DOI: 10.1016/S0166-1280(01)00802-8

ELSEVIER

Journal of Molecular Structure (Theochem) 583 (2002) 105–116

THEO CHEM

www.elsevier.com/locate/theochem

A structure–activity relationship (SAR) study of synthetic neolignans and related compounds with biological activity against *Escherichia coli*

A.J. Camargo[a], R. Mercadante[b], K.M. Honório[a], C.N. Alves[a], A.B.F. da Silva[a,*]

[a]*Departamento de Química e Fisica Molecular, Instituto de Química de São Carlos, Universidade de São Paulo, C.P. 780, 13560-970 São Carlos, SP, Brazil*

[b]*Departamento de Química, Universidade Estadual do Oeste do Paraná Campus de Toledo, Toledo, PR, Brazil*

Received 2 October 2001; accepted 28 November 2001

Abstract

Structure–activity relationship techniques were employed to classify neolignan compounds and related analogues against *Escherichia coli*. The AM1 (Austin Model) method was used to calculate a set of molecular descriptors (properties) for 16 synthetic neolignan compounds. The descriptors were further analyzed using the principal component analysis (PCA), hierarchical cluster analysis (HCA) and K-nearest neighbor (KNN) chemometric methods. The PCA and HCA methods showed quite efficient to classify the 16 compounds in two groups (active and inactive) and three descriptors were found to be important in the classification. By using the chemometric results, 14 new neolignan molecules were analyzed through the PCA, HCA and KNN methods and 11 of them are proposed as molecules potentially active against *E. coli*.

Keywords: *Escherichia coli*; Principal component analysis; Hierarchical cluster analysis; K-nearest neighbor

1. Introduction

Escherichia coli (*E. coli*) is a major component of the normal intestinal flora of humans and other mammals. A great diversity of commensal non-pathogenic *E. coli* strains belonging to many different serotypes can be isolated from the feces of healthy individuals. These strains are massively shed in the environment and may contaminate food of animal origin or other foods like vegetables, fruits and their derivatives. They may also contaminate surface and underground water, generally without any adverse effects on human health [1].

Some specific *E. coli* strains nevertheless represent primary pathogens with an enhanced potential to cause disease. These pathogens have broadly been classified into two major categories: the enteric pathogens and the extraintestinal pathogens. With the exception of Shiga toxin-producing *E. coli*, which also indirectly affect body parts other than the intestine, the enteric pathogens are agents of diarrhea in humans and animals. The extraintestinal pathogenic *E. coli* constitute a separate group mainly causing infections of the urinary tract in all age categories or sepsis and meningitis in small children and young animals [1].

* Corresponding author. Tel.: +55-162-739-964; fax: +55-162-739-975.

E-mail address: alberico@iqsc.sc.usp.br (A.B.F. da Silva).

0166-1280/02/$ - see front matter
PII: S0166-1280(01)00802-8

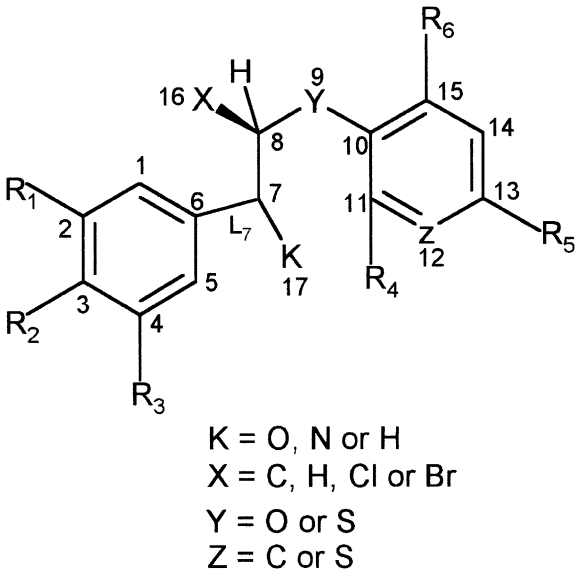

Fig. 1. General structure and numbering used in the neolignan compounds studied.

Due to the fact of this pathogen reach millions of people each year, in developed countries and in developing ones, there are many studies on the main factors responsible for the diseases caused by *E. coli*.

The aim of the present work is to investigate the relationship between some molecular properties and the activity against *E. coli* of synthetic neolignans and analogues by using quantum chemical and chemometric methods and afterwards to propose new neolignan analogs that can be potentially active against *E. coli*.

Neolignans are organic dimers derived from oxidative coupling of allyl and propenyl phenols that occur in the *Myristicaceae* [2–4] and other primitive plant families. The genus *Virola* presents about 35 species distributed in the neo-tropical, occurring mainly in the Amazonian forest [5–7]. The resin of several Virola supplies a hallucinogen powder used in ritual [8] and the genus *Virola* is a source of lignan and neolignan with recognized bioactivity [9].

In order to inhibit the metabolism process of *E. coli*, Ferri et al. [10] systematically synthesized a series of 16 neolignan analog compounds (Fig. 1 shows the general structure and numbering system used in this work). The evaluation of the toxicity against *E. coli* was done using the monitoring of CO_2, produced by the bacteria, during the breathing processes. The technique used for the CO_2 determination was the flow injection analysis with conductivity detection [10].

The tests showed that 13 out of the 16 neolignan compounds are inactive against *E. coli* while the other three compounds were considered active (Fig. 2). In this work we tried primarily to find out chemometric methods that could be able to classify the 16 compounds in two categories, actives and inactive, according to the calculated molecular properties of the compounds. Obtaining success in such classification, we would be able to propose a new group of possible active and inactive compounds to be synthesized and tested in laboratory.

2. Methodology

Neolignans are molecules that present several degrees of rotation with the possibility of obtaining many geometric conformations. In the absence of crystallographic structures, it is necessary to carry out a conformational search for the finding of the lowest conformational energy. In this work, we used a careful procedure for obtaining the lowest energy structure. Initially, a conformational search was carried out and the method employed was the Tripos 5.2 force field [11]. This search was performed with the software Spartan 5.0 [12] running on an Origin 2000 workstation. Next, the final molecular conformation of each compound studied was attained by using the AM1 semi-empirical method [13], as implemented in the molecular package AMPAC 5.0 [14], running on an Ultrasparc Sun Solaris workstation. 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.

Molecular properties of compounds are usually correlated with biological activity and this correlation study is known as structure–activity relationship (SAR) [15–17]. The SAR methods have been used successfully in pharmaceutical applications [18] 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 [19];

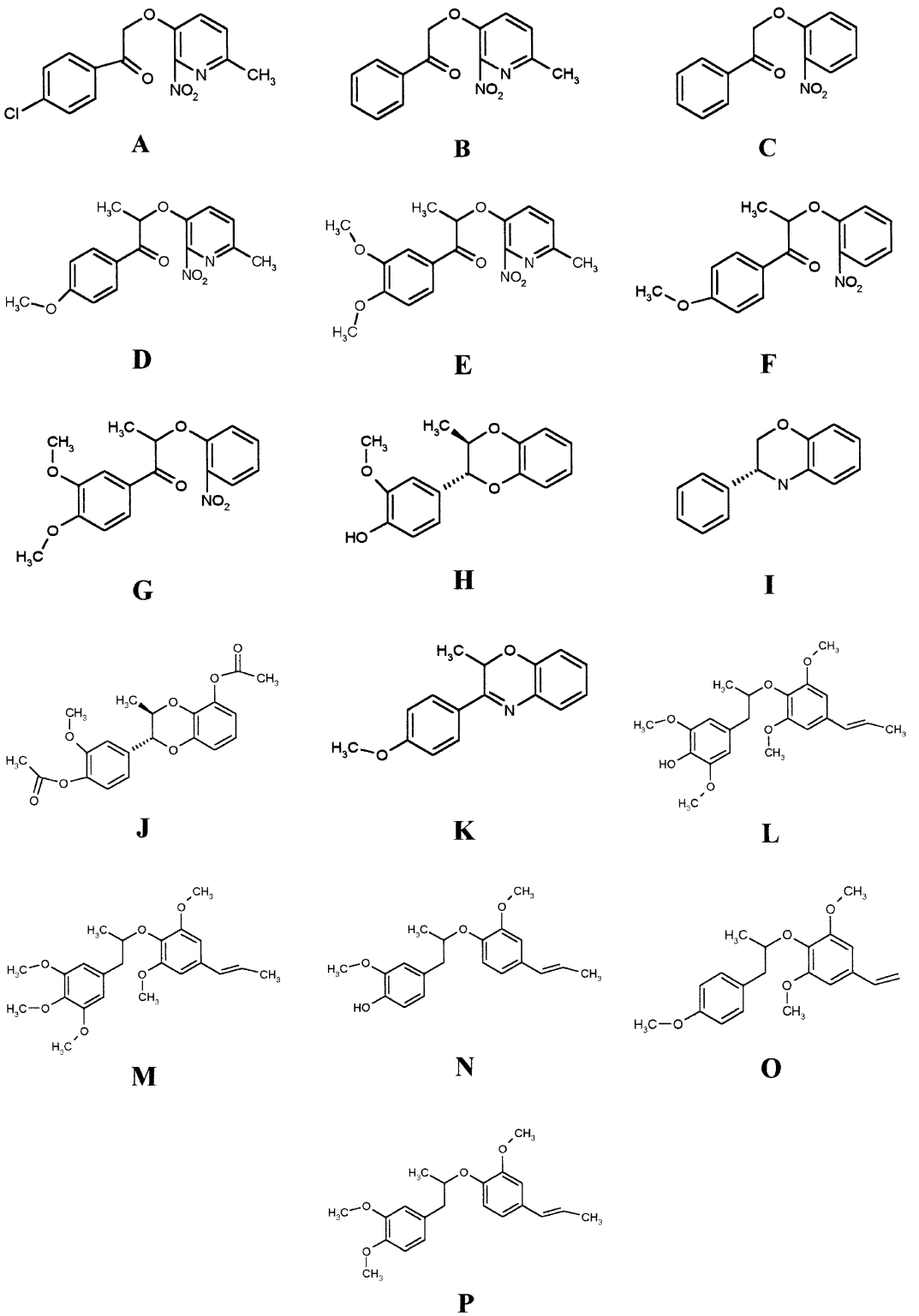

Fig. 2. Active (A, B and C) and inactive (D, E, F, G, H, I, J, K, L, M, N, O and P) neolignan compounds against *E. coli*.

- superficial area (A) and molecular volume (V): properties evaluated with the molecular package HyperChem 5.0 [20];
- partial atomic charges (Q_n) and bond orders (L_n): derived from the NBO analysis [21,22];
- energy of the HOMO (H) and LUMO (L) frontier orbitals;
- hardness (η): obtained from the equation $\eta = (E_{LUMO} - E_{HOMO})/2$;
- Mulliken electronegativity (χ): calculated from the equation $\chi = -(E_{HOMO} + E_{LUMO})/2$;
- other electronic properties were calculated: total energy (E_T), heat of formation (ΔH_f), ionization potential (PI), dipole moment (μ) and polarizability (pol), whose values were obtained from the molecular package Ampac 5.0 [14];
- dihedral (D_n) and interatomic (A_n) angles: obtained from the molecular package Ampac 5.0.

The correlation between molecular properties and biological activity was done by using the pattern recognition methods described in the Sections 2.1, 2.2 and 2.3 and built-in the computational package Pirouette [23].

2.1. Principal component analysis

Principal component analysis (PCA) is a multivariate statistical technique and the central idea of PCA is to reduce the dimensionality of a data set that presents a large number of interrelated variables, while retaining as much as possible the variation present in the data set. This is achieved by transforming to a new set of variables the principal components (PCs), which are uncorrelated, and which are ordered so that the first PCs retain most of the variation present in all of the original variables [24]. Three general aspects can be attained when the PCA analysis is employed:

- the original set of variables can be reduced to a smaller set that accounts for most of the variance in the data;
- PCA can search the data for qualitative and quantitative distinctions in situations where the number of data available is too large;
- PCA can be used to test hypotheses about the qualitative and quantitative distinctions in the data [25].

2.2. Hierarchical cluster analysis

The hierarchical cluster analysis (HCA) 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 useful for examining data sets for expected or unexpected clusters, including the presence of outliers. 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 the points belong to an only group [26].

2.3. K-nearest neighbor analysis [25]

The K-nearest neighbor (KNN) method classifies a new object (compound) according to its distance to an object of the training set. The closer neighbors of the training set are found and the object will be assigned into the class that have 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 others in the set but not with itself. The best value of K can be chosen based on the results from the training set alone [27]. 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.

3. Results and discussion

3.1. Pre-processing of the molecular descriptors

Before applying the pattern recognition methods to the 16 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 weighted equally and this provides a measure of the ability of a descriptor to discriminate classes of compounds [27]. With this method, we can compare all variables at the same level although presenting different units.

Table 1
Values obtained for the variance and Fisher's weights for all calculated variables

Variable	$W_{Fischer}$	W_{Var}	Variable	$W_{Fischer}$	W_{Var}	Variable	$W_{Fischer}$	W_{Var}	Variable	$W_{Fischer}$	W_{Var}
Q_1	0.265	2.530	Q_{22}	2.010	6.019	L_{12}	0.118	2.237	L_{33}	10.304	22.608
Q_2	0.043	2.085	Q_{23}	1.931	5.862	L_{13}	0.365	2.731	ΔH_f	1.613	5.225
Q_3	4.095	10.189	Q_{24}	1.028	4.056	L_{14}	0.729	4.359	E_T	0.421	2.842
Q_4	0.456	2.911	Q_{25}	0.622	3.244	L_{15}	1.261	4.523	PI	7.977	17.955
Q_5	0.843	3.685	Q_{26}	0.622	3.243	L_{16}	0.082	2.165	μ	1.743	5.485
Q_6	0.987	3.973	Q_{27}	5.955	13.911	L_{17}	0.765	3.530	Pol	1.754	5.509
Q_7	1.876	5.751	Q_{28}	0.032	2.064	L_{18}	0.857	3.714	A	1.270	4.540
Q_8	14.664	31.323	Q_{29}	0.639	3.278	L_{19}	0.103	2.207	V	1.427	4.854
Q_9	1.631	5.261	Q_{30}	0.625	3.249	L_{20}	0.057	2.114	H	7.972	17.944
Q_{10}	0.540	3.079	Q_{31}	11.718	25.437	L_{21}	0.625	3.249	L	1.575	5.149
Q_{11}	0.276	2.553	L_1	1.222	4.443	L_{22}	0.004	2.008	Log P	2.029	6.059
Q_{12}	0.601	3.202	L_2	1.275	4.550	L_{23}	2.306	6.611	D_1	0.443	2.886
Q_{13}	0.003	2.005	L_3	2.275	6.551	L_{24}	1.476	4.953	D_2	0.410	2.820
Q_{14}	0.183	2.366	L_4	0.036	2.072	L_{25}	0.444	2.889	D_3	0.371	2.743
Q_{15}	0.832	3.664	L_5	0.076	2.152	L_{26}	1.351	4.703	D_4	0.041	2.081
Q_{16}	0.828	3.656	L_6	1.267	4.535	L_{27}	0.571	3.142	A_1	0.057	2.114
Q_{17}	0.003	2.005	L_7	2.580	7.160	L_{28}	0.625	3.249	A_2	0.043	2.086
Q_{18}	0.061	2.122	L_8	0.836	3.671	L_{29}	2.619	7.239	A_3	0.934	3.868
Q_{19}	1.097	4.195	L_9	2.267	6.535	L_{30}	11.939	25.879	χ	4.940	11.808
Q_{20}	0.254	2.508	L_{10}	0.545	3.170	L_{31}	0.562	3.125	η	3.089	8.178
Q_{21}	2.014	6.028	L_{11}	0.096	2.192	L_{32}	0.625	3.250			

3.2. Choosing the molecular descriptors

After the autoscaling, it was evaluated the procedure for choosing the more relevant molecular descriptors and for that we used the correlation matrix between the calculated variables and the variance and the Fisher's weights [28,29]. This procedure gives the relative importance of each variable. The MATLAB program [30] was used for calculating the variance and Fisher's weights. The values of the variance and Fisher's weights are presented in Table 1.

From Table 1 we selected 10 variables that showed significant values of these weights, i.e. the variables that presented values above 3.00 for the Fisher's weight and values above 8.00 for the variance weight are those that possess a higher ability in the discrimination (separation) between active and inactive molecules.

From the 10 variables selected by variance and Fisher's weights, we obtained the correlation matrix between these variables and the respective calculated values. This correlation matrix is presented in Table 2

Table 2
Correlation matrix between the selected variables

	Q_3	Q_8	Q_{27}	Q_{31}	L_{30}	L_{33}	PI	H	χ	η
Q_3	1.00	0.64	−0.91	−0.90	0.91	0.81	−0.39	0.38	−0.32	−0.37
Q_8		1.00	−0.84	−0.87	0.85	0.92	−0.91	0.91	−0.87	−0.54
Q_{27}			1.00	0.98	−0.97	−0.97	0.63	−0.63	0.62	0.35
Q_{31}				1.00	−0.99	−0.96	0.66	−0.67	0.63	0.42
L_{30}					1.00	0.95	−0.64	0.64	−0.60	−0.42
L_{33}						1.00	−0.75	0.75	−0.76	−0.36
PI							1.00	−1.00	0.95	0.63
H								1.00	−0.95	−0.63
χ									1.00	0.36
η										1.00

Table 3
Values obtained for the three most important properties (descriptors) that classified the 16 compounds of the training set as active and inactive molecules

Compound	χ (eV)	η (eV)	Q_3
A	5.451	4.456	−0.033
B	5.572	4.580	−0.096
C	5.305	4.433	−0.096
D	5.227	4.276	0.114
E	5.008	4.038	0.093
F	5.234	4.392	0.114
G	5.008	3.995	0.091
H	4.284	4.437	0.069
I	4.731	4.068	−0.115
J	4.356	4.422	0.039
K	4.625	3.967	0.097
L	4.388	4.116	0.043
M	4.443	4.227	0.007
N	4.388	4.192	0.056
O	4.371	4.243	0.074
P	4.343	4.222	0.051

and from it we can see that some variables are correlated to each other (we considered correlated variables those that possess correlation coefficients above 0.70) and, according to the results showed in Tables 1 and 2, only five variables can be considered important for the separation between active and inactive compounds.

3.3. Statistical analysis

After several analysis, the best separation was obtained by using the following variables: χ, η and Q_3, whose values are presented in Table 3. This suggests 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 56.63% of the variance of the data. Considering the first (PC1) and second (PC2) components, the accumulated variance increases to 79.44%. Fig. 3 and Table 4 show that PC1 is in fact responsible for the discrimination

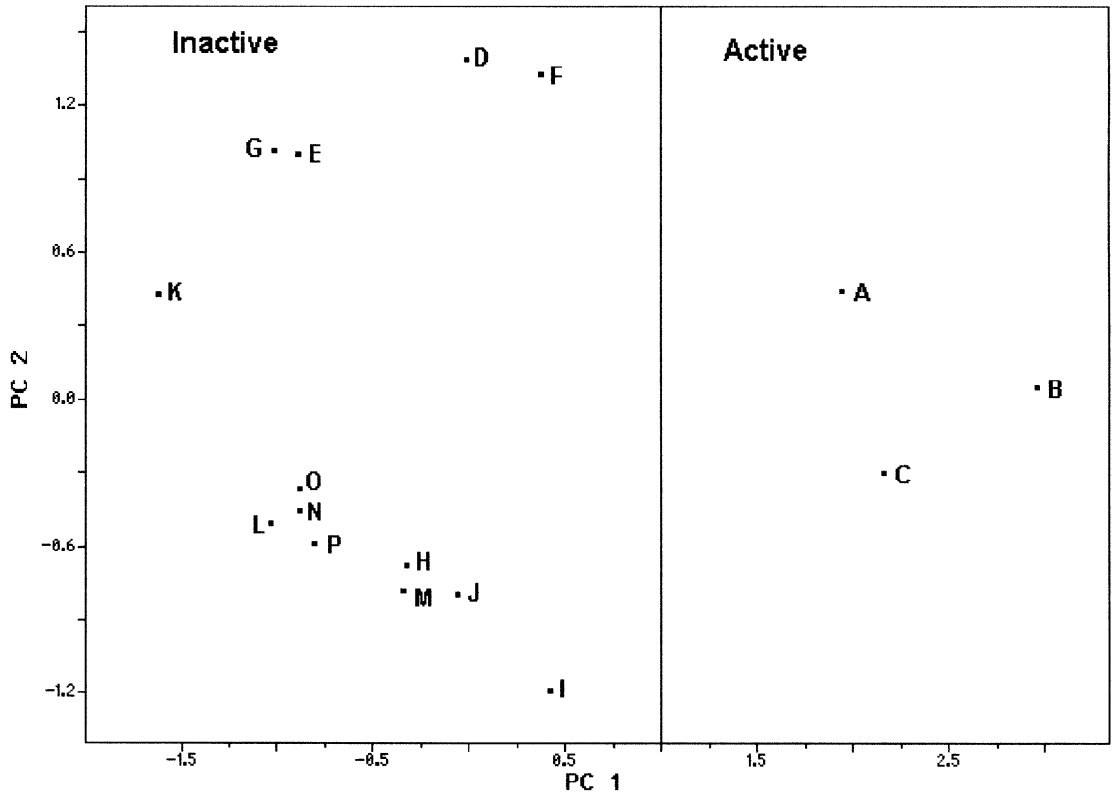

Fig. 3. The separation of the training set into two groups: active and inactive compounds. Notice that the first component is responsible for the separation.

Table 4
PCA scores obtained for the 16 compounds studied with the two principal components

Compound	PC1	PC2
A	1.945	0.446
B	2.964	0.054
C	2.161	−0.302
D	−0.001	1.389
E	−0.887	1.002
F	0.371	1.325
G	−1.010	1.013
H	−0.328	−0.680
I	0.421	−1.191
J	−0.062	−0.795
K	−1.620	0.433
L	−1.033	−0.508
M	−0.339	−0.782
N	−0.886	−0.451
O	−0.879	−0.366
P	−0.809	−0.588

Table 5
PCA loadings obtained for the selected variables with the two principal components

Variable	PC1	PC2
χ	0.563	0.768
η	0.594	−0.121
Q_3	−0.574	0.629

between active (A, B and C) and inactive (D, E, F, G, H, I, J, K, L, M, N, O and P) compounds. From Fig. 4 and Table 5 we can see that the variables χ, η and Q_3 have high PC1 loadings and PC1 can be represented by the following equation:

$$\mathrm{PC1} = 0.563[\chi] + 0.594[\eta] - 0.574[Q_3] \quad (1)$$

Fig. 3 shows the separation of the training set of molecules into two groups: active and inactive molecules against *E. coli* when we used the variables χ, η and Q_3 to obtain the separation. From Fig. 3 we can see that

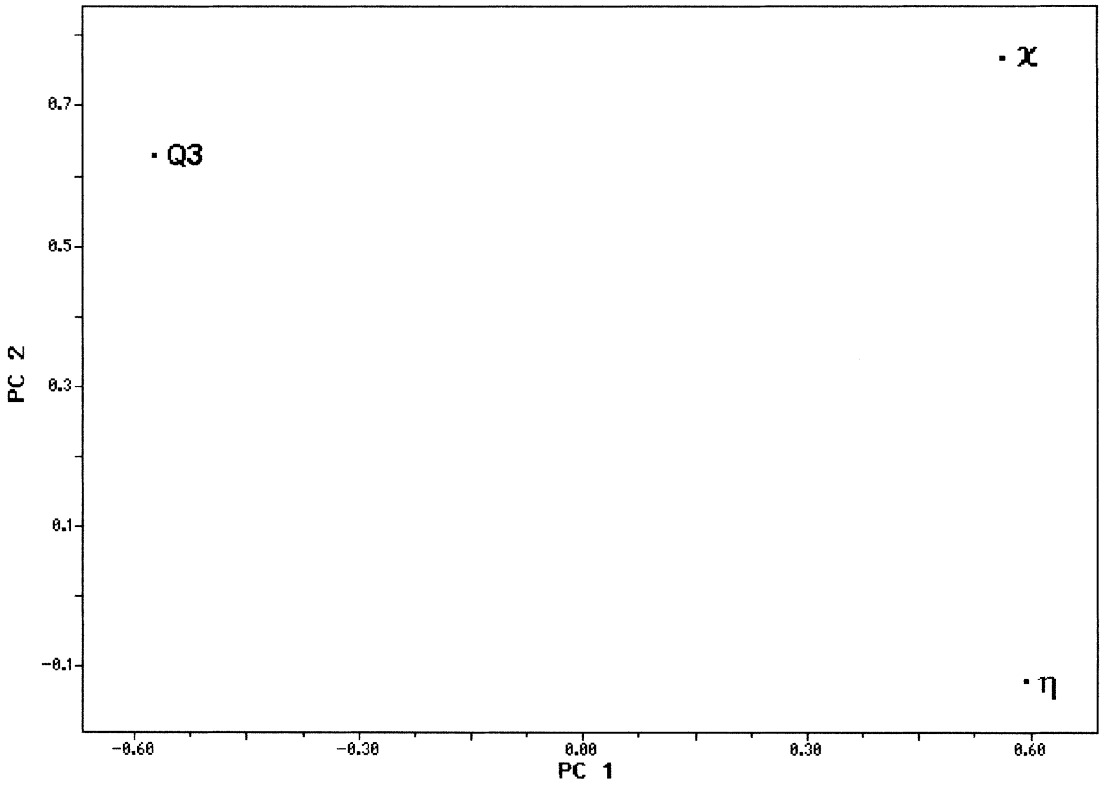

Fig. 4. Plot of the loading values of the selected variables used in the training set.

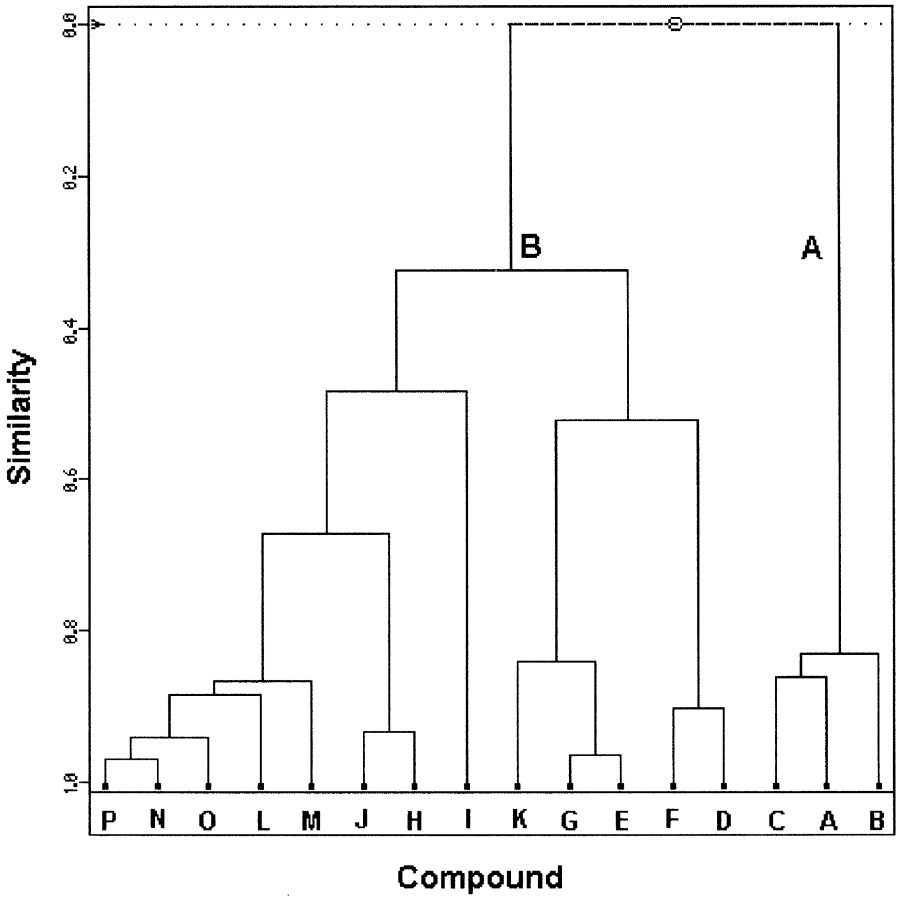

Fig. 5. Dendrogram obtained with the HCA method. The training set was classified into two groups: A (active) and B (inactive).

the active compounds present a very positive values for PC1 while the inactive compounds present a progressive decrease in the PC1 values.

From Fig. 4 and Eq. (1) we can see that for a compound to become active it needs to present large and positive values for the electronegativity (χ) and molecular hardness (η) along with small values for the charge on atom 3 (Q_3).

It is also interesting to notice that the variables responsible for the separation between active and inactive compounds, i.e. χ, η and Q_3, are all electronic variables, therefore we can conclude that electronic effects have a very important role when one is trying to understand the activity of neolignan compounds against *E. coli*.

The electronegativity (χ) determines the direction of electron transfer between two chemicals when they react [31], thus for our active compounds we can conclude that the greater the electronegativity value, the greater the possibility of charge transfer reactions occur between the compounds studied and the biological receptor.

Regarding to absolute hardness (η), Pearson brought forth an unifying concept which enabled the rationalization of chemical reactivities, selectivities and stabilities of compounds [32]. Chemical entities including molecules were categorized as 'hard' and 'soft' Lewis acids or bases (the 'hard' species in general having small atomic radius, high effective nuclear charge and low polarizability) [32]. Pearson also formulated the 'hard and soft acids and bases principle' [32] and this principle states that acids show greater affinity for bases of the same class and vice versa. Thus, hard acids (acceptors) tend to form

Table 6
Classification obtained with the KNN method

Category	Number of compounds	Compounds incorrectly classified	
		1NN	3NN
Active	3	0	0
Inactive	13	0	0
Total	16	0	0
Percentage of correct information		100	100

strong bonds with hard bases (donors), but bind reluctantly or weakly to soft bases and the latter class of compounds interacts preferably with soft acids; in other words, a hard–soft combination is destabilized. From our results, the active compounds must present high hardness, i.e. the increase of the value of absolute hardness lead to an increase in the activity against *E. coli* and this is related to the higher rate of reaction with the biological receptor of the same class (hard).

On the charge at atom 3 we would like to pay attention to the fact that for the active compounds it is important to have atoms with small charge at position 3, i.e. more electron-acceptor atoms at position 3 are required for the active compounds, as they would react with an electron-donor biological receptor.

The HCA results are abridged in the dendrogram showed in Fig. 5. From this dendrogram we can notice that the similarity observed between the group of the active (A) and inactive (B) molecules is zero.

The HCA and PCA methods are complementary and for the 16 neolignan compounds studied in this work the HCA and PCA results were similar, i.e. HCA and PCA classified the 16 neolignan compounds under study exactly in the same way.

The KNN method was used for the validation of the initial data set and Table 6 presents the results obtained with one (1NN) and three (3NN) nearest neighbors. For both cases (1NN and 3NN) the percentage of correct information was 100% and we decided to use 3NN instead of 1NN because the greater the number of nearest neighbors, the better the reliability of the KNN method.

The KNN results were also similar to those obtained with the HCA and PCA methods. The outcomes obtained with the three classification methods, PCA, HCA and KNN, were quite interesting, as not very often is possible to obtain 100% of success in classifying a data set using these three methods.

After the employing of the three classification methods, we decided to apply them to a series of new neolignan compounds whose activity against *E. coli* was not yet known. Fig. 6 shows the chemical structure of the 14 new neolignan compounds, numbered from E1 to E14, specially designed to present activity against *E. coli*. These compounds are similar to those used in the training set (the initial 16 compounds). The calculated molecular properties for the 14 new compounds are showed in Table 7. In order to verify if these new molecules would be active or inactive, we had to apply the results obtained with the three pattern recognition methods, PCA, HCA and KNN, previously employed in our 16 molecules.

Table 8 shows the results of the PCA calculations for the first (PC1) and second (PC2) principal components. Before carrying out the prediction calculations, the variables were also autoscaled as previously. Three compounds (E2, E9 and E10) were classified as inactive by PCA and KNN and all the other ones were classified as active for both methods. The HCA method classified only two compounds (E9 and E10) as being inactive and all the other compounds as being active. All the three results are summarized in Table 9. From Table 9 we can conclude that the compounds classified as active by the three methods (E1, E3, E4, E5, E6, E7, E8, E11, E12, E13 and E14) are certainly potentially active against *E. coli*. Thus, these compounds are suggested for synthesis and future biological tests against *E. coli* with neolignan molecules.

4. Conclusions

The application of three pattern recognition methods (PCA, HCA and KNN) in neolignan compounds with activity against *E. coli* 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 χ, η and

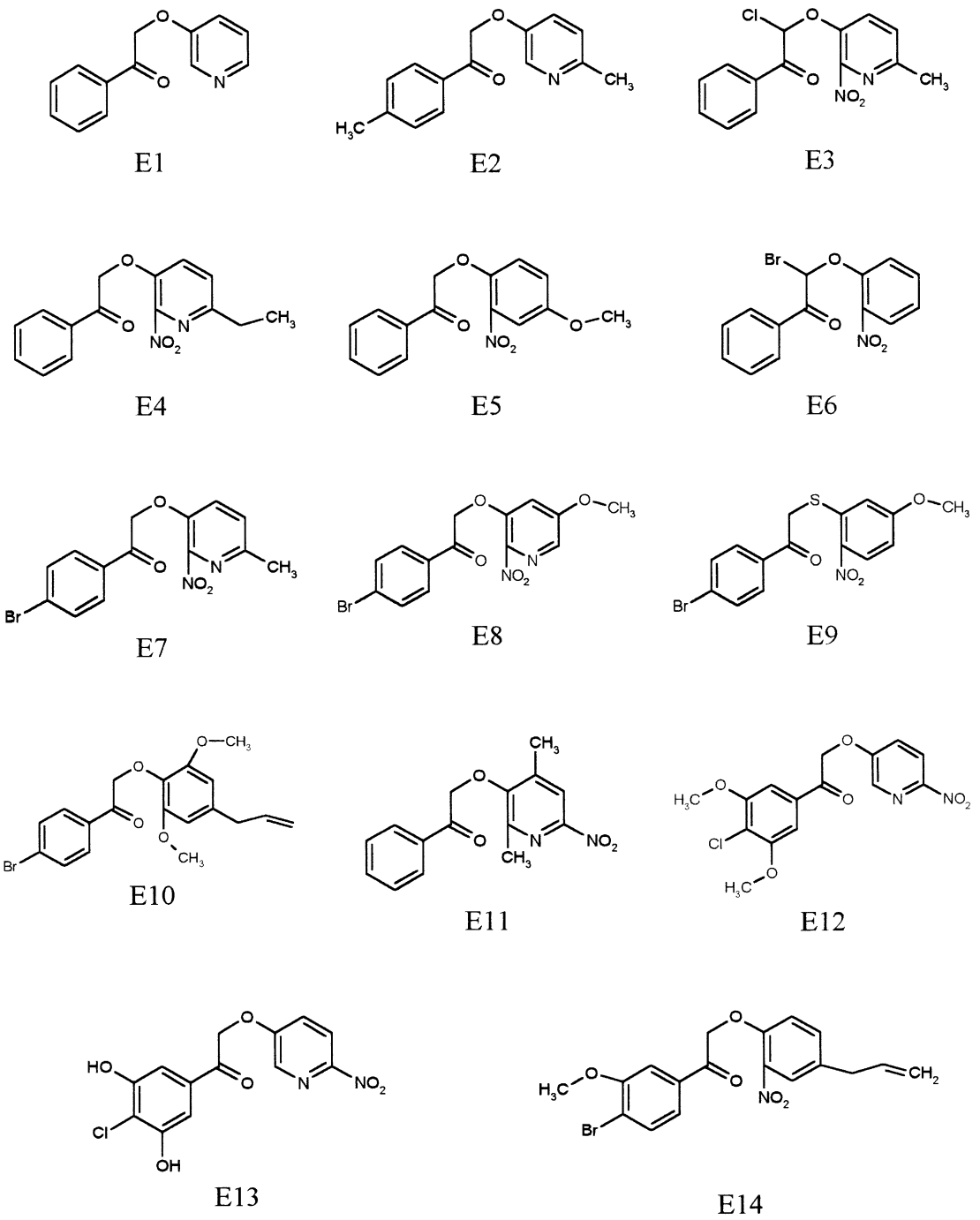

Fig. 6. Compounds selected for the activity prediction against *E. coli*.

Table 7
Theoretical descriptors calculated for the new set of neolignan analog compounds

Compound	χ (eV)	η (eV)	Q_3
E1	4.936	4.353	−0.098
E2	5.006	4.184	−0.035
E3	5.651	4.411	−0.095
E4	5.328	4.539	−0.097
E5	5.144	4.166	−0.098
E6	5.592	4.499	−0.103
E7	5.516	4.367	−0.140
E8	5.378	4.439	−0.141
E9	4.902	3.845	−0.139
E10	4.612	3.955	−0.143
E11	5.584	4.460	−0.096
E12	5.462	4.159	−0.136
E13	5.614	4.269	−0.157
E14	5.292	4.356	−0.156

Q_3 are responsible for the separation between active and inactive compounds. The HCA results were similar to those obtained with PCA, i.e. both methods classified the 16 neolignan compounds under study exactly in same way.

The results obtained for the training set (16 neolignan compounds) by using the three methods (PCA, HCA and KNN) were applied for the test set (14 new analog compounds) in order to verify if these new compounds would be active or inactive. The results showed that 11 compounds are potentially active against *E. coli*.

Table 8
Compound scores from Table 7

Compound	PC1	PC2
E1	0.445	−0.030
E2	−0.261	−0.330
E3	1.597	0.134
E4	1.478	0.689
E5	0.261	−0.825
E6	1.787	0.452
E7	1.556	−0.317
E8	1.550	−0.001
E9	−0.670	−2.374
E10	−0.777	−1.909
E11	1.634	0.337
E12	0.920	−1.139
E13	1.544	−0.841
E14	1.302	−0.435

Table 9
The results obtained with the three pattern recognition methods, PCA, HCA and KNN, for the 14 new compounds: active compound (+) and inactive compound (−)

Compound	PCA	HCA	KNN
E1	+	+	+
E2	−	+	−
E3	+	+	+
E4	+	+	+
E5	+	+	+
E6	+	+	+
E7	+	+	+
E8	+	+	+
E9	−	−	−
E10	−	−	−
E11	+	+	+
E12	+	+	+
E13	+	+	+
E14	+	+	+

Acknowledgements

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

References

[1] P. Kuhnert, P. Boerlin, J. Frey, FEMS Microbiol. Rev. 24 (2000) 107.
[2] M. Yoshida, O.R. Gottlieb. Proceedings of the Brazilian Symposium on the Chemistry of Lignins and Other Wood Components, São Carlos, Brazil (1989) p. 301.
[3] O.R. Gottlieb, Phytochemistry 11 (1942) 1537.
[4] O.R. Gottlieb, M. Yoshida, Química Nova 7 (1984) 250.
[5] O.R. Gottlieb, Fortschr. Chem. Org. Naturst. 35 (1977) 1.
[6] D.A. Whiting, Nat. Prod. Rep. 7 (1990) 349.
[7] W.A. Rodrigues, Acta Amazonica 10 (1980) 1.
[8] E.W. Davis, J.A. Yost, J. Ethnopharmacol. 9 (1983) 273.
[9] W.D. Macrae, G.H.N. Towers, Phytochemistry 23 (1984) 1207.
[10] P.H. Ferri, L.E.S. Barata, Phytochemistry 31 (1992) 1375.
[11] M. Clark, R.D. Cramer, N.V. Opdensch, J. Comput. Chem. 10 (1989) 982.
[12] SPARTAN 5.0, Wavefunction, Irvine, CA, 1997.
[13] M.J. Dewar, E.G. Zoebisch, E.F. Healy, J.J.P. Stewart, J. Am. Chem. Soc. 107 (1985) 3902.
[14] AMPAC 5.0, Semichem., 7128, Shawnee, KS 66216, 1994.
[15] W. Karcher, J. Devillers (Eds.), Chemical and Environmental Science, vol. 1, Kluwer Academic Publishers, Dordrecht, 1990.
[16] C.N. Alves, L.S. Santos, I.N. Jardim, J. Braz. Chem. Soc. 9 (1998) 574.

[17] C.N. Alves, L.G.M. de Macedo, K.M. Honório, A.J. Camargo, L.S. Santos, I.N. Jardim, L.E.S. Barata, A.B.F. da Silva, J. Braz. Chem. Soc. (2002) in press.
[18] H. Kubinyi, QSAR: Hansch Analysis and Related Approaches, VCH, New York, 1993.
[19] G.G. Nys, R.F. Rekker, Eur. J. Med. Chem. Chim. Ther. 9 (1974) 361.
[20] HyperChem, Release 4.5 for Windows, Reference Manual, Canada, 1995.
[21] J.E. Carpenter, F. Weinhold, J. Mol. Struct. (Theochem) 169 (1988) 41.
[22] J.E. Carpenter. PhD Thesis, University of Wisconsin, Madison, WI, 1987.
[23] Pirouette Multivariate Data Analysis for IBM PC Systems, version 2.0. Infometrix: Seattle, WA, 1996.
[24] I.T. Jolliffe, Principal Component Analysis, Springer, New York, 1986.
[25] W.R. Dillon, M. Goldstein, Multivariate Analysis, Wiley, New York, 1984.
[26] B.R. Kowalski, C.F. Bender, J. Am. Chem. Soc. 94 (1972) 5632.
[27] J.C. Lindon, E. Holmes, J.K. Nicholson, Prog. Nuc. Res. Spec. 39 (2001) 1.
[28] R.E. Bruns, J.F.G. Faigle, Química Nova 84 (1985) 84.
[29] M.A. Sharaf, D.L. Illman, B.R. Kowalski, Chemometrics, Wiley, New York, 1986.
[30] MATLAB 4.2c.1 for Windows, the Mathworks, 1994.
[31] R.G. Pearson, J. Am. Chem. Soc. 107 (1985) 6801.
[32] R.G. Pearson, J. Am. Chem. Soc. 85 (1963) 3533.