Internet Electronic Journal of Molecular Design **2004**, *3*, 759–770 ISSN 1538–6414
BioChem Press http://www.biochempress.com

Inter*net* Electronic Journal of Molecular Design

December 2004, Volume 3, Number 12, Pages 759–770

Editor: Ovidiu Ivanciuc

Proceedings of the Internet Electronic Conference of Molecular Design, IECMD 2003
November 23 – December 6, 2003
Part 6

Use of an Expert System in Lignan Skeleton Prediction from ^{1}H NMR Data

Mara B. Costantin,[1] Marcelo J. P. Ferreira,[1] Gilberto V. Rodrigues,[2] and Vicente P. Emerenciano [1]

[1] Instituto de Química, Universidade de São Paulo, Caixa Postal 26077, 05513–970, São Paulo, SP, Brazil
[2] Departamento de Química, ICEx, Universidade Federal de Minas Gerais, 30161–000, Belo Horizonte, MG, Brazil

Received: November 1, 2003; Revised: August 2, 2004; Accepted: September 9, 2004; Published: December 31, 2004

Citation of the article:
M. B. Costantin, M. J. P. Ferreira, G. V. Rodrigues, and V. P. Emerenciano, Use of an Expert System in Lignan Skeleton Prediction from ^{1}H NMR Data, *Internet Electron. J. Mol. Des.* **2004**, *3*, 759–770, http://www.biochempress.com.

Inter*net* Electronic Journal of Molecular Design

BioChem Press
http://www.biochempress.com

Use of an Expert System in Lignan Skeleton Prediction from ^{1}H NMR Data [#]

Mara B. Costantin,[1] Marcelo J. P. Ferreira,[1] Gilberto V. Rodrigues,[2] and Vicente P. Emerenciano [1,*]

[1] Instituto de Química, Universidade de São Paulo, Caixa Postal 26077, 05513–970, São Paulo, SP, Brazil

[2] Departamento de Química, ICEx, Universidade Federal de Minas Gerais, 30161–000, Belo Horizonte, MG, Brazil

Received: November 1, 2003; Revised: August 2, 2004; Accepted: September 9, 2004; Published: December 31, 2004

Internet Electron. J. Mol. Des. **2004**, *3* (12), 759–770

Abstract

Motivation. The identification of natural products is a task that demands skilled spectroscopists in this area. The number of isolated substances has been enhanced significantly in the last years. Therefore, trying to reduce the time spent during the identification this process, we developed a program able to predict the carbon skeleton for a compound from ^{1}H NMR data. For that, the natural product class selected in this study was the lignans, products showing some complex structures.

Method. The H1MACH program was developed to assist the process of skeleton prediction of organic compounds from the ^{1}H NMR chemical shift data. Thus, a database containing 760 ^{1}H NMR spectra data was established. From the data, the program can predict the most probable skeleton type for a new compound under analysis and show several structures that have a high similarity index with the supplied data.

Results. The program was evaluated with 30 lignan structures not stored yet in the database. The results show that the program was able to predict, in 70% of the studied cases, the correct skeleton of the compound. Analyzing with more details the results, one can verify that in 90% of the tests, the correct skeleton was predicted among the three first skeletons by the program.

Conclusions. Regarding the obtained results, it can be concluded that the tests carried out with the program H1MACH showed good results, once that the signal multiplicity was not included in the database. The procedure here described can be applied for other classes of compounds. This new tool will increase the power of spectral data interpretation of the expert system SISTEMAT.

Keywords. Lignan; expert system; ^{1}H NMR; structure elucidation; skeleton prediction; computer–aided analysis.

Abbreviations and notations

^{13}C NMR, Carbon–13 nuclear magnetic resonance	^{1}H NMR, Proton nuclear magnetic resonance
SI, similarity index	

[#] Presented in part at the Internet Electronic Conference of Molecular Design, IECMD 2003.

* Correspondence author; phone: 55–11–3091–2056; fax: 55–11–3815–5579; E–mail: vdpemere@iq.usp.br.

1 INTRODUCTION

In the last decades, the discovery of a large number of compounds originated from plants has been intensified due to the development of new techniques of separation and isolation as well as to the modern apparatus related to their identification. However, the structure elucidation of compounds from NMR data by the natural products' researchers may be still slow as a consequence of the increasing diversity of classes and complexity of skeletons of such compounds. Our group has developed the expert system SISTEMAT [1,2], especially for the chemistry of natural products with the aim of helping the researchers during the structure elucidation process, enabling them to obtain the most likely skeletons of these compounds promptly and successfully. SISTEMAT is composed of various applicative programs, which are the intelligent part of the expert system since they permit the analyses of the data contained in the system and present useful information during the structural elucidation of an unknown compound [2]. Currently, the SISTEMAT database stores a huge number of spectral data obtained from several plant chemical classes, such as, monoterpenes, iridoids, lactonic and non–lactonic sesquiterpenes, diterpenes and triterpenes [3–8].

The present work introduces the first class of aromatic compounds, lignans, in the SISTEMAT. Lignans are derived from two phenylpropanoids units linked by a bond between C8 and C8' and consist of an important class of natural products mainly due to their physiological activity [9]. We also describe the use of H1MACH, an applicative program included in the SISTEMAT, which is responsible, in our case, for lignan skeleton prediction using ^{1}H NMR data collected from literature. In the case of lignans, ^{1}H NMR spectra were selected rather than ^{13}C NMR due to the facility in elucidating ^{1}H NMR spectra of aromatic compounds and also to the small sample quantity and little time required for analyses in relation to ^{13}C NMR. Consequently, the amount of ^{1}H NMR spectra contained in literature becomes superior to the amount of ^{13}C NMR one. The analyses of 30 lignans carried out in this study show a very good performance of the H1MACH with 70% accuracy.

2 MATERIALS AND METHODS

For the compilation of the database information, the lignans having ^{1}H NMR or systematic data were collected from the literature (1970–2002). The database of this study contains 800 lignans distributed in 40 different skeletons, 760 ^{1}H NMR spectra data and 113 occurrences in 30 families of plants.

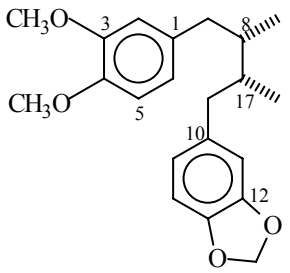
2.1 The Applicative Programs

2.1.1 The DATASIS program

The ^{1}H NMR spectra of lignans were inserted in the DATASIS program [10], which encodes automatically the molecular structural drawing done by the user since the compound should be

recognized by the microcomputer in a mathematical language, in a process called molecular codification [1]. This codification method also allows the system to find out chemical information of the compound from the encoded structure.

DATASIS also includes a database with the chemical structural of the compounds in addition with their data collected from the literature, *i.e.*, class and skeleton of the molecule, family, genus and species of the plant, physico–chemical data and bibliography. After the drawing and codification of each structure, the ^{1}H NMR data of the compounds were stored into database. Like the signal multiplicity is extremely affected by the spectrometer type, 200, 300, 400 or 500MHz, and the data encountered in the literature show the most varied types of chemical shifts, the multiplicity standard has not been established and included in the database. Thus, only the ^{1}H NMR chemical shifts were used.



Manglietiastrum sinicum – Magnoliaceae
Data from ^{1}H NMR spectra: ($CDCl_3$) H2 6.63, H5 6.77, H6 6.67, H7 2.73 / 2.25, H8 1.71, H9 0.81, H11 6.64, H14 6.70, H15 6.59, H16 2.69 / 2.28, H17 1.73, H18 0.83, OCH_2O 5.89

Figure 1. Lignan used to exemplify the program H1MACH.

2.1.2 The H1MACH program

After the total insertion of ^{1}H NMR data of the lignans in DATASIS, the structural determination of a lignan compound is processed by H1MACH. This program matches the ^{1}H NMR spectral data of a compound with all data stored in the database and attributes for each compound a similarity percent (SI) with the test sample. This percent is computed with the Bremser's system [11]. After this attribution, the program selects the *x*–lignans that exhibit the higher similarity percentual with the tested sample and shows them to the user. The *x*–value is selected by the user in the initial analysis and may vary from 1 to 25.

Table 1. Skeleton probability shows by the program H1MACH

Skeleton	Probability
06	74.90
09	24.67
19	0.43

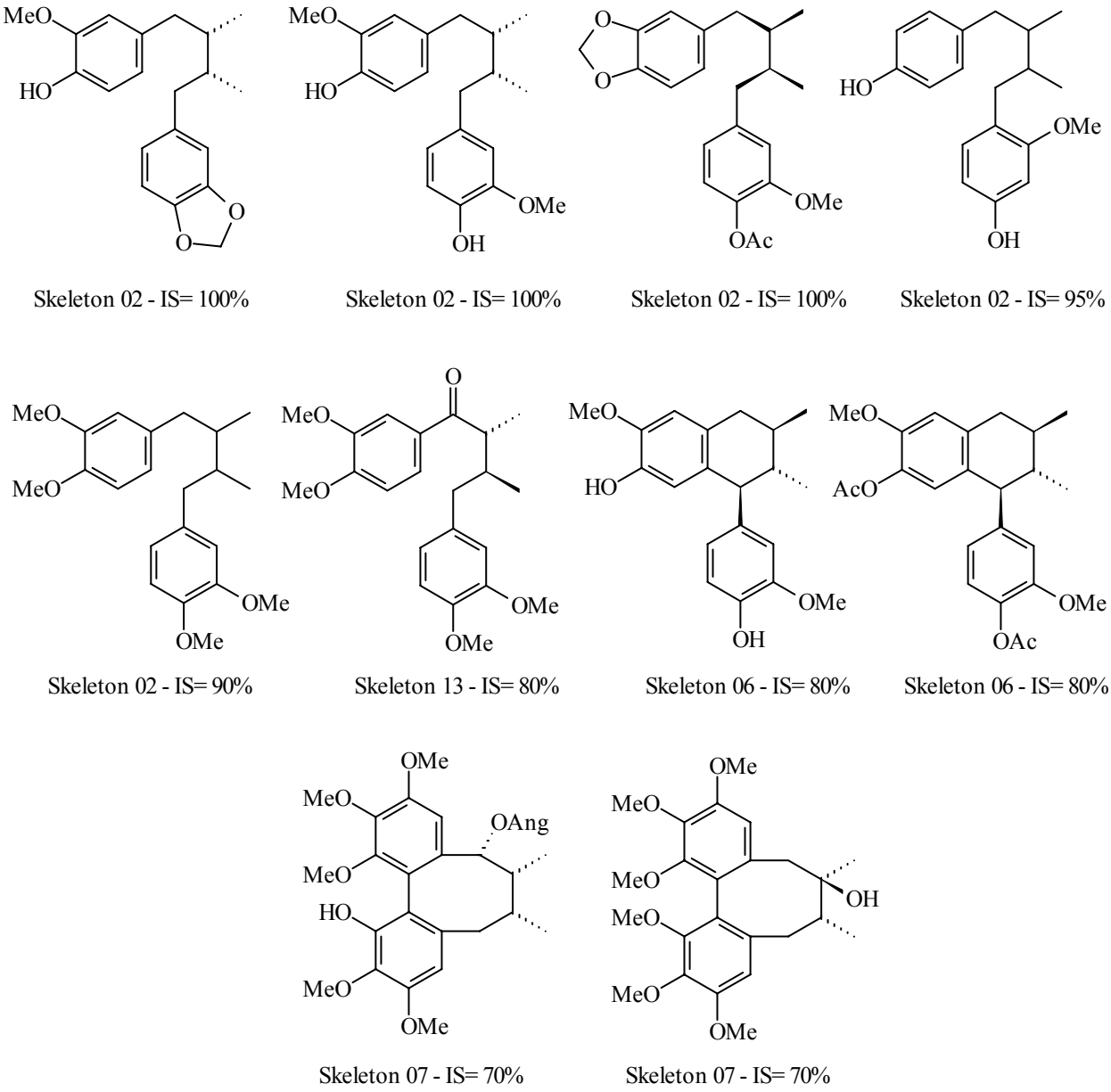


Figure 2. Substances with higher SI exhibited by the H1MACH program.

The lignan isolated from *Manglietiastrum sinicum* (Magnoliaceae) [12], Figure 1, was tested with the purpose of exemplifying the applying of the H1MACH program. Figure 1 shows the lignan and the respective ^{1}H NMR literature data. After the input data, the program carries out the data match and shows the ten compounds with the highest similarity index, at a chemical shift range of 0.5δ, which are shown in Figure 2. At the final step of analysis, the program exhibits the skeleton probability for the compound tested. This probability (P) is calculated by: P= (NS/TNS)×100, where NS is the number of times that a determined skeleton was found, and TNS is the total number of selected substances. Thus, the skeleton probability of the substance in question is computed. Table 1 exhibits the skeleton probability furnished by the H1MACH.

3 RESULTS AND DISCUSSION

To test the performance and efficiency of the program H1MACH, our option was to randomly collect the ^{1}H NMR spectra data of 30 lignans (Figure 3) from the literature published recently [13–33], which were not inserted into our database. The results obtained with the program H1MACH are shown in Table 2. This table also exhibits the first three lignan skeletons suggested by the program, their ^{1}H NMR data and respective references. The skeletons proposed by the program are presented in Figure 4.

Higher specialization of spectroscopists in the identification of natural product classes jointly with the great diversity and structural complexity of the skeletons stimulates the development of computational programs that assist in the skeleton prediction and structural elucidation of new compounds isolated from natural sources as well as in the identification of substances even reported in literature.

The results shown in Table 2 demonstrate that the program H1MACH predicts the correct skeleton in 70% of the lignans tested, indicating a very good accuracy, considering the higher similarity exhibited among the lignan skeletons and the diversity of substituent groups found in the compounds.

The negative results (compounds L–10, L–24 and L–26) presented by the H1MACH were due to the great existing similarity existent between the tested chemical shifts and the chemical shifts stored in the database, which pertain to the other skeletons. In these cases, the correct skeletons were not shown. For the compounds L–2, L–3, L–12, L–14, L–23 and L–25 the program displayed wrong results because the correct skeleton was not presented as the most probable, however the correct skeleton was always proposed as one the three first options. On the other hand, it is noteworthy to show here that for these compounds, except in test L–14, the first substance listed by the program was the compound that exhibited the higher similarity index with the tested experimental data and pertains to the correct skeleton of the substance. For all cases, including negative and positive cases, the similar skeletons were present in the database.

This study is one the first carried out by our research group that utilizes only ^{1}H NMR chemical shifts data for skeleton prediction of a determined class of natural products. So, this program in next future will be integrated in a set of programs that performs structural elucidation and skeleton prediction from the multispectral data.

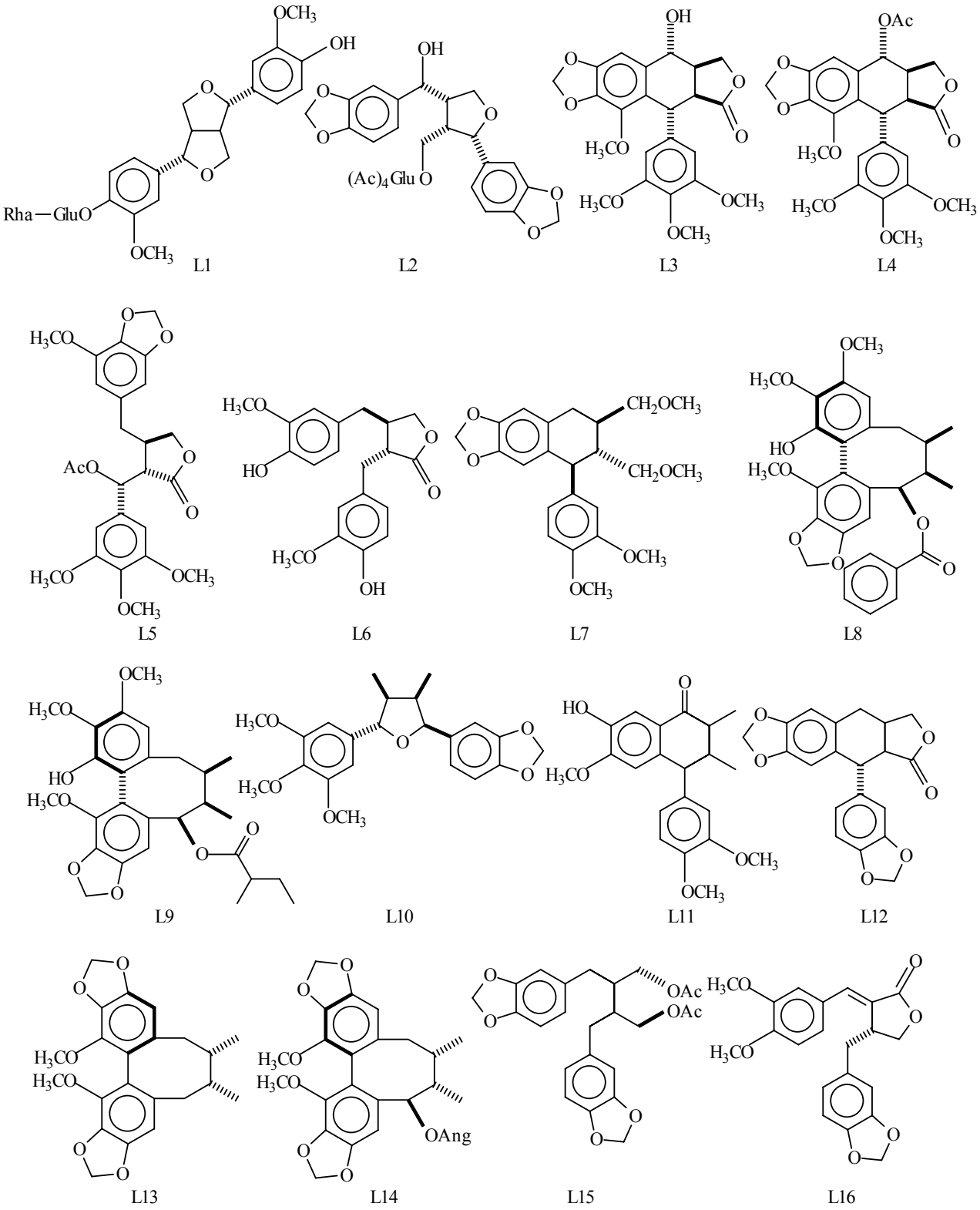


Figure 3. Lignans used to test the H1MACH program.

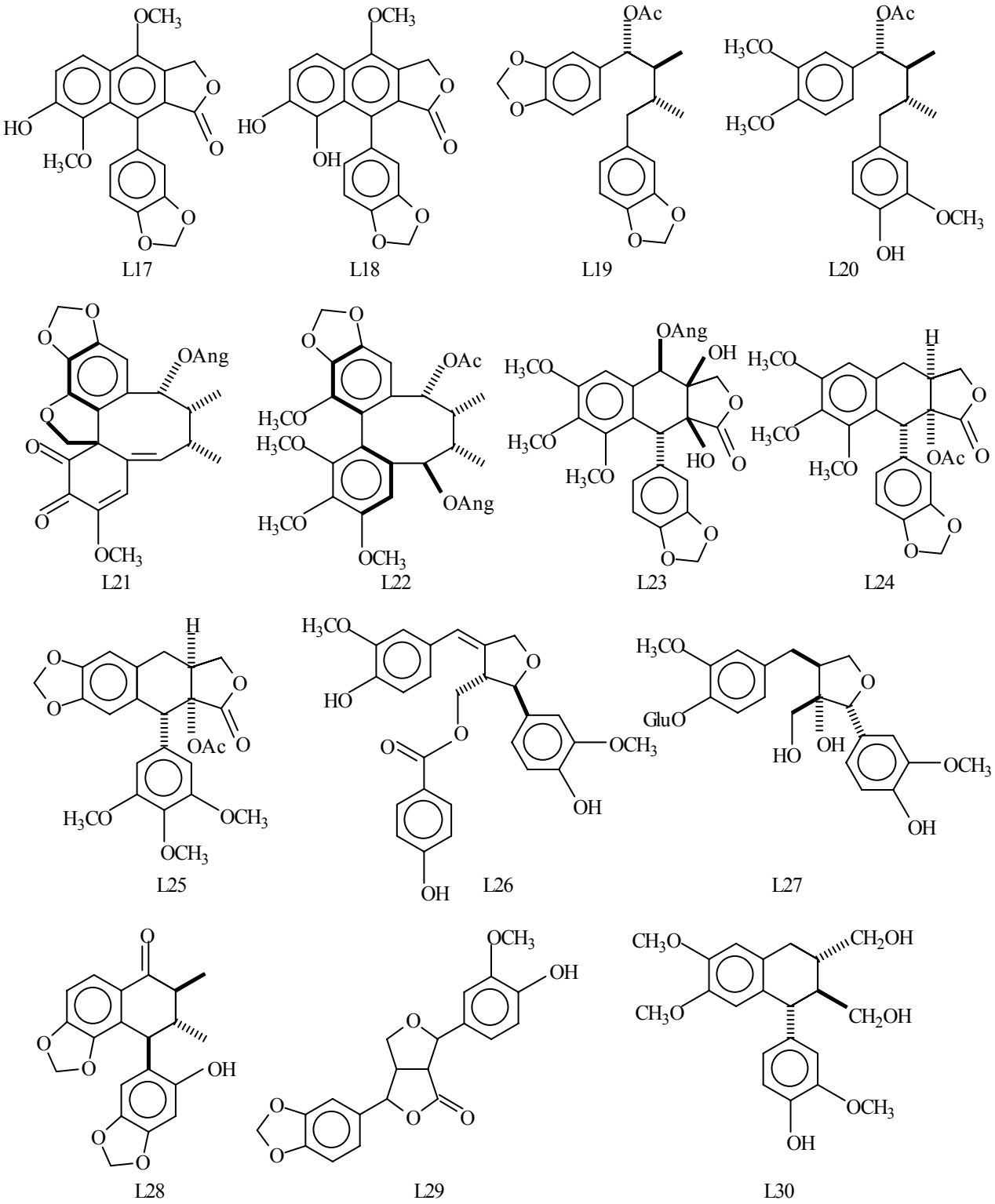


Figure 3. (Continued)

Table 2. Results obtained with the H1MACH program

Compound, skeleton and botanical data	^{1}H NMR literature data	Skeleton Probability*	Refs.
L–1, Ske 01 (Malvaceae) *Hibiscus syriacus*	DMSO–d_6: H2– 6.92, H5– 7.02, H6– 6.85, H7– 4.68, H8– 3.04, H9– 3.75 and 4.13, H11– 6.87, H14– 6.71, H15– 6.74, H16– 4.62, H17– 3.04, H18– 3.75 and 4.13; Glucose H1– 4.99, H2– 3.54, H3– 3.44, H4– 3.18, H5– 3.29, H6– 3.63 and 3.43; Rhamnose H1– 5.22, H2– 3.68, H3– 3.37, H4– 3.19, H5– 3.92, H6– 1.11	*Ske 01 – 100%*	[13]
L–2, Ske 03 (Scrophulariaceae) *Lancea tibetica*	DMSO–d_6: H2– 6.80, H5– 6.80, H6– 6.80, H7– 4.81, H8– 2.31, H9– 3.99 and 3.59, H11– 6.80, H14– 6.80, H15– 6.80, H16– 5.58, H17– 2.85, H18– 3.44 and 3.58; OCH_2O: 5.96, 5.97; Glucose H1– 4.24, H2– 3.02, H3– 3.18, H4– 3.09, H5– 3.12, H6– 3.51 and 3.69	Ske 02 – 63.43% Ske 01 – 31.06% *Ske 03 – 5.20%*	[14]
L–3, Ske 05 (Hernandiaceae) *Hernandia sonora*	$CDCl_3$: H2– 6.83, H7– 4.74, H8– 2.68, H9– 4.05 and 4.51, H11– 6.35, H15– 6.35, H16– 4.61, H17– 2.68; OCH_2O: 5.88, 5.90	Ske 04 – 62.98% Ske 15 – 0.63% *Ske 05 – 35.76%*	[15]
L–4, Ske 05 (Hernandiaceae) *Hernandia sonora*	$CDCl_3$: H2– 6.53, H7– 5.87, H8– 2.81, H9– 4.19 and 4.87, H11– 6.38, H15– 6.38, H16– 4.36, H17– 2.80; OCH_2O: 5.86, 5.91	*Ske 05 – 98.78%* Ske 04 – 1.22%	[15]
L–5, Ske 04 (Hernandiaceae) *Hernandia sonora*	$CDCl_3$: H2– 6.04, H6– 5.97, H7– 2.29 and 2.46, H8– 2.78, H9– 3.96 and 4.31, H11– 6.39, H15– 6.39, H16– 6.14, H17– 2.75; OCH_2O: 5.91, 5.94	*Ske 04 – 63.90%* Ske 06 – 36.10%	[15]
L–6, Ske 04 (Selaginellaceae) *Selaginella doederleinii*	$CDCl_3$: H2– 6.41, H5– 6.51, H6– 6.80, H7– 2.52, H8– 2.52, H11– 6.59, H14– 6.59, H15– 6.82, H16– 2.91, H17– 2.52, H18– 3.87, 4.14; OH: 2 X 5.48	*Ske 04 – 99.67%* Ske 08 – 0.17% Ske 09 – 0.17%	[16]
L–7, Ske 06 (Euphorbiaceae) *Phyllanthus niruri*	$CDCl_3$: H2– 6.18, H5– 6.56, H7– 2.79, H8– 1.78, H9– 3.25, H11– 6.59, H14– 6.79, H15– 6.67, H16– 3.91, H17– 2.14, H18– 3.25; OCH_2O: 5.80, 5.81	*Ske 06 – 70.50%* Ske 02 – 29.50%	[17]
L–8, Ske 07 (Schisandraceae) *Kadsura longipedunculata*	$CDCl_3$: H2– 6.58, H7– 2.72, H9– 1.20, H11– 6.44, H16– 5.82, H18– 1.08; OCH_2O: 5.90, 5.94	*Ske 07 – 99.0%* Ske 04 – 1.0%	[18]
L–9, Ske 07 (Schisandraceae) *Kadsura longipedunculata*	$CDCl_3$: H2– 6.40, H7– 2.66, H9– 0.82, H11– 6.52, H16– 5.61, H18– 1.08; OCH_2O: 5.94, 5.98	*Ske 07 – 99.0%* Ske 04 – 1.0%	[18]
L–10, Ske 10 (Schisandraceae) *Schisandra henryi*	$CDCl_3$: H2– 6.61, H6– 6.61, H7– 4.63, H8– 2.45, H9– 1.03, H11– 6.85, H14– 6.85, H15– 6.85, H16– 5.43, H17– 2.45, H18– 0.63; OCH_2O: 5.95	Ske 02 – 74.99% Ske 06 – 23.44% Ske 04 – 0.84%	[19]
L–11, Ske 06 (Schisandraceae) *Schisandra sphenanthera*	$CDCl_3$: H2– 7.64, H5– 6.44, H8– 2.76, H9– 1.10, H11– 6.60, H14– 6.80, H15– 6.53, H16– 3.97, H17– 2.76, H18– 0.98; OH: 5.60	*Ske 06 – 62.36%* Ske 02 – 35.09% Ske 11 – 1.60%	[20]
L–12, Ske 05 (Burseraceae) *Commiphora incisa*	$CDCl_3$: H2– 6.64, H5– 6.58, H7– 2.46 and 2.81, H8– 3.04, H9– 3.96 and 4.44, H11– 6.60, H14– 6.60, H15– 6.73, H16– 4.36, H17– 3.31; OCH_2O: 5.92, OCH_2O: 5.93	Ske 04 – 88.99% *Ske 05 – 11.01%*	[21]
L–13, Ske 07 (Schisandraceae) *Kadsura longipedunculata*	$CDCl_3$: H2– 6.52, H9– 0.75, H11– 6.52, H16– 2.10, H18– 0.98; OCH_2O: 5.97, OCH_2O: 5.98	*Ske 07 – 69.23%* Ske 12 – 15.38% Ske 08 – 7.70%	[22]

Table 2. (Continued)

Compound, skeleton and botanical data	^{1}H NMR literature data	Skeleton Probability*	Refs.
L–14, Ske 07 (Schisandraceae) *Kadsura longipedunculata*	$CDCl_3$: H2– 6.41, H9– 0.94, H11– 6.70, H16– 5.70, H18– 0.83; OCH_2O: 5.96, OCH_2O: 6.00, 6.02	Ske 02 – 81.30% *Ske 07 – 13.01%* Ske 12 – 3.25%	[22]
L–15, Ske 02 (Burseraceae) *Bursera ariensis*	$CDCl_3$: H2– 6.55, H5– 6.68, H6– 6.51, H7– 2.60, H8– 2.10, H9– 4.06, H11– 6.55, H14– 6.68, H15– 6.51, H16– 2.60, H17– 2.10, H18– 4.06; OCH_2O: 5.88, OCH_2O: 5.88	*Ske 02 – 71.28%* Ske 04 – 28.72%	[23]
L–16, Ske 04 (Apiaceae) *Chaerophyllum maculatum*	$CDCl_3$: 2.80 (2H), 3.84 (1H), 4.24 (2H), 6.66 (2H), 6.70 (1H), 6.94 (1H), 7.06 (1H), 7.22 (1H), 7.52 (1H); OCH_2O: 5.88	*Ske 04 – 100.0%*	[24]
L–17, Ske 13 (Acanthaceae) *Justicia flava*	$CDCl_3$/CD_3OD: H2– 7.95, H3– 7.37, H9– 5.53, H11– 7.01, H14– 7.01, H15– 7.01; OCH_2O: 5.86	*Ske 13 – 90.74%* Ske 06 – 7.41% Ske 14 – 1.85%	[25]
L–18, Ske 13 (Acanthaceae) *Justicia flava*	$CDCl_3$/CD_3OD: H2– 7.98, H3– 7.40, H9– 5.52, H11– 6.80, H14– 6.80, H15– 6.80; OCH_2O: 5.80	*Ske 13 – 96.97%* Ske 06 – 1.52% Ske 14 – 1.52%	[25]
L–19, Ske 02 (Magnoliaceae) *Talauma ovata*	$CDCl_3$: H2– 6.54, H5– 6.54, H6– 6.54, H7– 5.52, H8– 1.57, H9– 0.91, H11– 6.54, H14– 6.54, H15– 6.54, H16– 2.40, H17– 1.57, H18– 0.78; OCH_2O: 5.94, 5.90	*Ske 02 – 83.58%* Ske 07 – 16.42%	[26]
L–20, Ske 02 (Magnoliaceae) *Talauma ovata*	$CDCl_3$: H2– 6.61, H5– 6.61, H6– 6.61, H7– 5.53, H8– 1.71, H9– 0.94, H11– 6.61, H14– 6.61, H15– 6.61, H16– 2.40, H17– 1.71, H18– 0.83; OH: 5.47	*Ske 02 – 81.97%* Ske 04 – 14.26% Ske 07 – 3.78%	[26]
L–21, Ske 15 (Schisandraceae) *Kadsura interior*	$CDCl_3$: H2– 6.50, H9– 5.82, H8– 1.88, H9– 1.04, H11– 6.67, H16– 5.83, H17– 3.03, H18– 1.02, H19– 4.40 and 4.84; OCH_2O: 6.02	*Ske 15 – 59.58%* Ske 07 – 40.42%	[27]
L–22, Ske 07 (Schisandraceae) *Kadsura interior*	$CDCl_3$: H2– 6.44, H7– 5.73, H8– 2.22, H9– 1.03, H11– 6.71, H16– 5.84, H17– 2.12, H18– 0.94; OCH_2O: 5.95	*Ske 07 – 79.71%* Ske 12 – 11.59% Ske 4, 8, 16–2.90%	[27]
L–23, Ske 05 (Burseraceae) *Commiphora erlangeriana*	$CDCl_3$: H2– 6.98, H7– 5.76, H9– 4.35 and 3.57, H11– 6.53, H14– 6.68, H15– 6.49, H16– 4.95; OCH_2O: 5.91	Ske 04 – 52.49% Ske 01 – 45.25% *Ske 05 – 1.81%*	[28]
L–24, Ske 05 (Burseraceae) *Commiphora erlangeriana*	$CDCl_3$: H2– 6.55, H7– 2.45 and 3.33, H8– 3.05, H9– 3.79 and 4.76, H11– 6.81, H14– 6.70, H15– 6.69, H16– 5.12; OCH_2O: 5.91	Ske 06 – 49.84% Ske 01 – 49.84% Ske 03 – 0.16%	[28]
L–25, Ske 05 (Burseraceae) *Commiphora erlangeriana*	$CDCl_3$: H2– 6.72, H5– 6.61, H7– 2.50 and 3.26, H8– 2.99, H9– 3.87 and 4.74, H11– 6.49, H15– 6.49, H16– 4.35; OCH_2O: 5.90, 5.92	Ske 06 – 60.61% *Ske 05 – 38.79%* Ske 04 – 0.61%	[28]
L–26, Ske 03 (Apiaceae) *Agastache rugosa*	CD_3OD: H2– 6.94, H5– 6.74, H6– 6.82, H7– 5.14, H8– 3.80, H9– 4.20 and 4.70, H11– 7.05, H14– 6.80, H15– 6.97, H16– 6.50, H18– 4.63	Ske 01 – 88.89% Ske 10 – 11.11%	[29]
L–27, Ske 03 (Valerianaceae) *Valeriana officinalis*	CD_3OD: H2– 6.92, H5– 6.73, H6– 6.73, H7– 4.82, H8– 3.80, H9– 3.60 and 3.79, H11– 6.90, H14– 7.09, H15– 6.77, H16– 2.54 and 3.12, H17– 2.59, H18– 3.63 and 4.05	*Ske 03 – 100.0%*	[30]

Table 2. (Continued)

Compound, skeleton and botanical data	^{1}H NMR literature data	Skeleton Probability*	Refs.
L–28, Ske 06 (Myristicaceae) *Virola sebifera*	$CDCl_3$: H2– 7.62, H3– 6.73, H8– 2.36, H9– 1.17, H12– 6.34, H15– 6.33, H16– 4.12, H17– 1.99, H18– 0.96; OCH_2O: 5.84, OCH_2O: 5.67, 5.77	*Ske 06 – 66.88%* Ske 01 – 32.02% Ske 07 – 0.50%	[31]
L–29, Ske 01 (Styracaceae) *Styrax officinalis*	$CDCl_3$: H2– 6.72, H5– 6.72, H6– 6.72, H7– 5.25, H8– 3.10, H9– 4.10 and 4.50, H11– 6.85, H14– 6.85, H15– 6.85, H16– 5.25, H17– 3.35; OCH_2O: 5.95	Ske 04 – 39.30% *Ske 01 – 59.16%* Ske 03 – 0.79%	[32]
L–30, Ske 06 *Scaphopetalum thonneri*	CD_3OD: H2– 6.28, H5– 6.71, H7– 2.80 and 2.81, H8– 2.01, H9– 3.68, H11– 6.69, H14– 6.76, H15– 6.63, H16– 3.88, H17– 1.78, H18– 3.42 and 3.71	Ske 02 – 22.86% *Ske 06 – 48.57%* Ske 04 – 28.57%	[33]

* The skeletons that are in italics represent the correct skeleton of the lignan

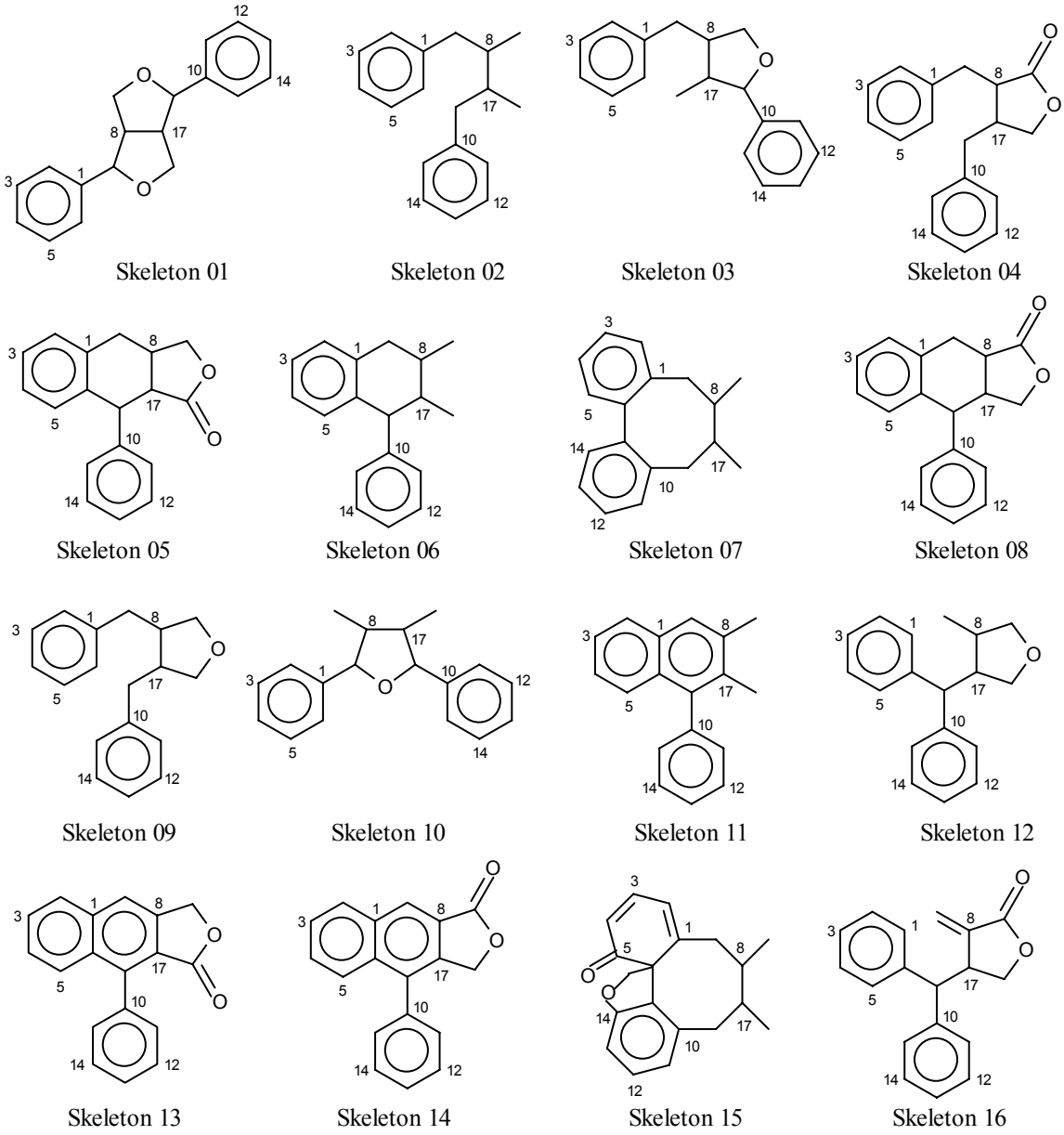


Figure 4. Skeletons proposed by the H1MACH program.

4 CONCLUSIONS

This study is one the first one carried out by our research group, that utilized only ^{1}H NMR chemical shifts data for skeleton prediction of a determined class of natural products. Regarding the obtained results, it can be concluded that the tests performed by the program H1MACH showed good results, once that the signal multiplicity was not included in the database. For the cases where the program mistakes the skeleton prediction, it was observed that the correct skeleton of the substance is found among the three first skeletons proposed by the program in 90.0% of the cases. Thus, one can affirm that the program H1MACH successfully carried out the identification of lignan skeletons. The structural elucidation might be more efficient with the introduction of other data, such as, ^{13}C NMR and natural sources. So, in future, this program will be integrated in a set of programs that perform structural elucidation from the multispectral data.

Acknowledgment

This work was supported by grants from the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) and by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). The authors thank Antônio J.C. Brant for helpful discussion during the preparation of the manuscript.

5 REFERENCES

[1] J. P. Gastmans, M. Furlan, M. N. Lopes, J. H. G. Borges and V. P. Emerenciano, A inteligência artificial aplicada à química de produtos naturais. O programa SISTEMAT. Parte I – Bases Teóricas, *Quím. Nova* **1990**, *13*, 10–16.

[2] J. P. Gastmans, M. Furlan, M. N. Lopes, J. H. G. Borges and V. P. Emerenciano, A inteligência artificial aplicada à química de produtos naturais. O programa SISTEMAT. Parte II – Organização do programa e aplicativos, *Quím. Nova* **1990**, *13*, 75–80.

[3] M. J. P. Ferreira, V. P. Emerenciano, G. A. R. Linia, P. Romoff, P. A. T. Macari and G. V. Rodrigues, ^{13}C NMR spectroscopy of monoterpenoids, *Prog. Nucl. Magn. Res. Spec.* **1998**, *33*, 153–206.

[4] M. J. P. Ferreira, G. V. Rodrigues, V. P. Emerenciano, MONOREG – an expert system for structural elucidation of monoterpenes, *Can. J. Chem.* **2001**, *79*, 1915–1925.

[5] G. V. Rodrigues, Um sistema especialista em determinação estrutural de sesquiterpenos lactonizados. Um enfoque multiespectral aliado a dados botânicos, PhD. Thesis, IQ–USP, **1996**.

[6] F. C. Oliveira, M. J. P. Ferreira, C. V. Núñez, G. V. Rodriguez and V. P. Emerenciano, ^{13}C NMR spectroscopy of eudesmane sesquiterpenes, *Prog. Nucl. Magn. Res. Spec.* **2000**, *37*, 1–45.

[7] S. A. V. Alvarenga, J. P. Gastmans, G. V. Rodrigues and V. P. Emerenciano, Ditregra – an auxiliary program for structural determination of diterpenes, *Spectroscopy*. **1997**, *13*, 227–249.

[8] P. A. T. Macari, J. P. Gastmans, G. V. Rodrigues and V. P. Emerenciano, An Expert system for Structure Elucidation of Triterpenes, *Spectroscopy* **1994/1995**, *2*, 139–167.

[9] G. P. Moss, Nomenclature of lignans and neolignans, *Pure Appl. Chem.* **2000**, *72*, 1493–1523.

[10] V. P. Emerenciano, G. V. Rodrigues and J. P. Gastmans, Um método de codificação semi–automático assistido por computador para o SISTEMAT, *Quím. Nova*. **1993**, *16*, 431–434.

[11] W. Bremser, M. Klier and E. Meyer, Mutual assignment of subspectra and substructures – A way to structure elucidation by ^{13}C NMR spectroscopy, *Org. Magn. Res.* **1975**, *7*, 97–137.

[12] B–G. Wang, X. Hong, L. Li, J. Zhou and X–J. Hao, Chemical constituents of two Chinese Magnoliaceae plants, *Tsoongiodendron odorum* and *Manglietiastrum sinicum*, and their inhibition of platelet aggregation. *Planta Med.* **2000**, *66*, 511–515.

[13] S–J. Lee, Y–S. Yun, I–K. Lee, I–J Ryoo, B–S. Yun and I–D. Yoo, An Antioxidant Lignan and other Constituents from the Root Bark of *Hibiscus syriacus*. *Planta Med*. **1999**, *65*, 658–660.

[14] B. Su, Q. Zhu, K. Gao, C. Yuan and Z. Jia Lignan and Phenylpropanoid Glycosides from *Lancea tibetica* and their Antitumor Activity. *Planta Med*. **1999**, *65*, 558–561.

[15] L. Udino, J. Abaul, P. Bourgeois, L. Gorrichon, H. Duran and C. Zedde, Lignans from the Seeds of *Hernandia sonora*, *Planta Med*. **1999**, *65*, 279–281.

[16] R. C. Lin, A–L. Skaltsounis, E. Seguin, F. Tillequin and M. Koch, Phenolic Constituents of *Selaginella doederleinii*, *Planta Med*.. **1994**, *60*, 168–170.

[17] Y. L. Huang, C. C. Chen and J. C. Ou, Isolintetralin: a New Lignan from *Phyllanthus niruri*, *Planta Med*. **1992**, *58*, 473–474.

[18] L–N. Li, H. Xue and X. Li, Three New Dibenzocyclooctadiene Lignans from *Kadsura longipedunculata*, *Planta Med*. **1991**, *57*, 169–171.

[19] L–L. Niang and X. Hong, Henricine, a New Tetrahydrofuran Lignan from *Schisandra henryi*, *Planta Med*. **1986**, *52*, 493–494.

[20] L–L. Niang and X. Hong, Schisandrone, a New 4–aryltetralone Lignan from *Schisandra sphenanthera*, *Planta Med*. **1985**, 51, 217–219.

[21] G.J. Provan and P.G. Waterman, Picropolygamain: a New Lignan from *Commiphora incisa* resin, *Planta Med*. **1985**, 51, 271–272.

[22] T. Rui, L. L. Niang and F. Qicheng, Studies on the Chemical Constituents of *Kadsura longipedunculata*: Isolation and Structure Elucidation of Five New Lignans, *Planta Med*. **1984**, 50, 414–417.

[23] J. D. Hernández, L. U. Román, J. Espiñeira and P. J–Nathan, Ariensin, a New Lignan from *Bursera ariensis*, *Planta Med*. **1983**, *47*, 215–217.

[24] G. A. Mikaya, D. G. Turabelidze, E. P. Kemertelidze and N. S. Wulfson, Kaerophyllin, A New Lignan from *Chaerophyllum maculatum*, *Planta Med*. **1981**, *43*, 378–380.

[25] A. A. Olaniyi, Two New Arylnaphthalide Lignans from *Justicia flava* Roots, *Planta Med*. **1982**, *44*, 154–156.

[26] M. E. A. Stefanello, M. A. Alvarenga, I. N. Toma, New Neolignans from *Talauma ovata*, *Fitoterapia* **2002**, *73*, 135–139.

[27] D. F. Chen, S. X. Zhang, M. Kozuka, Q. Z. Sun, J. Feng, Q. Wang, T. Mukainaka, Y. Nobukuni, H. Tokuda, H. Nishino, H. K. Wang, S. L. M. Natschke and K. H. Lee, Interiotherins C and D, Two New Lignans from *Kadsura interior* and Antitumor–Promoting Effects of Related Neolignans on Epstein–Barr Virus Activation, *J. Nat. Prod*. **2002**, *65*, 1242–1245.

[28] A. Dekebo, M. Lang, K. Polborn, E. Dagne, W. Steglich, Four lignans from *Commiphora erlangeriana*, *J. Nat. Prod*. **2002**, *65*, 1252–1257.

[29] C. Lee, H. Kim, Y. Kho, Agastinol and Agastenol, Novel Lignans from *Agastache rugosa* and Their Evaluation in an Apoptosis Inhibition Assay, *J. Nat. Prod*. **2002**, *65*, 414–416.

[30] B. Schumacher, S. Scholle, J. Hölzl, N. Khudeir, S. Hess and C. E. Müller, Lignans Isolated from Valerian: Identification and Characterization of a New Olivil Derivative With Partial Agonistic Activity at A_1 Adenosine Receptors, *J. Nat. Prod*. **2002**, *65*, 1479–1485.

[31] K. R. Rezende and M. J. Kato Dibenzylbutane and Aryltetralone Lignans from Seeds of *Virola sebifera*, *Phytochemistry* **2002**, *61*, 427–432.

[32] A. Ulubelen, Y. Saiki, H. Lotter, V. M. Chari and H. Wagner, Chemical Components of *Styrax officinalis* L. –IV The Structure of a New Lignan Styraxin, *Planta Med*. **1978**, *34*, 403–407.

[33] J. C. Vardamides, A. G. B. Azabaze, A. E. Nkengfack, F. R. van Heerden, Z. T. Fomum, T. M. Ngando, J. Conrad, B. Vogler and W. Kraus, Scaphopetalone and scaphopetalumate, a lignan and a tripterpene ester from *Sacaphopetalum thonneri*, *Phytochemistry* **2003**, *62*, 647–650.