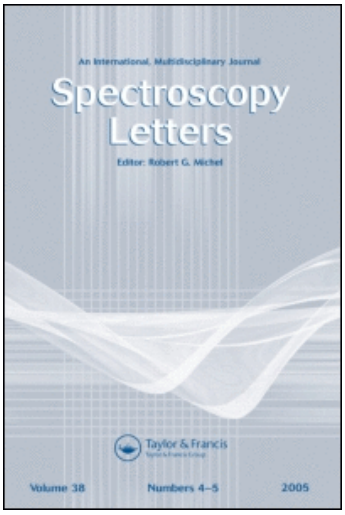


Spectroscopy Letters
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713597299

Computer-Assisted Approach to Structural Elucidation of Lignans

Mara B. Costantin[a]; Marcelo J. P. Ferreira[a]; Gilberto V. Rodrigues[b]; Vicente P. Emerenciano[a]
[a] Instituto de Química, Universidade de São Paulo, São Paulo, SP, Brazil [b] Departamento de Química, ICEx, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brazil

To cite this Article Costantin, Mara B. , Ferreira, Marcelo J. P. , Rodrigues, Gilberto V. and Emerenciano, Vicente P.(2008) 'Computer-Assisted Approach to Structural Elucidation of Lignans', Spectroscopy Letters, 41: 8, 405 — 421
To link to this Article: DOI: 10.1080/00387010802371817
URL: http://dx.doi.org/10.1080/00387010802371817

Spectroscopy Letters, 41:405–421, 2008

ISSN: 0038-7010 print/1532-2289 online
DOI: 10.1080/00387010802371817

Taylor & Francis
Taylor & Francis Group

Computer-Assisted Approach to Structural Elucidation of Lignans

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, São Paulo, SP, Brazil
[2]Departamento de Química, ICEx, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brazil

ABSTRACT This paper reports an expert system (SISTEMAT) developed for structural determination of diverse chemical classes of natural products, including lignans, based mainly on ^{13}C NMR and ^{1}H NMR data of these compounds. The system is composed of five programs that analyze specific data of a lignan and shows a skeleton probability for the compound. At the end of analyses, the results are grouped, the global probability is computed, and the most probable skeleton is exhibited to the user. SISTEMAT was able to properly predict the skeletons of 80% of the 30 lignans tested, demonstrating its advantage during the structural elucidation course in a short period of time.

KEYWORDS computer-aided analyses, lignans, NMR, structure elucidation

INTRODUCTION

Lignans comprise an important class of natural products derived from two phenylpropanoid units bonded between C-8 and C-8′.[1] Lignoids attract attention of researchers due to their large occurrence in nature and the variety of skeletons and biological activities.[2–4] Lignans, like podophyllotoxin, often have interesting pharmacological properties. In the 1940s, the investigation of the chemistry and pharmacology of *Podophyllum peltatum* root extracts led to the isolation of this lignan and the demonstration that it was a powerful inhibitor of mitosis. Inhibitors of cell division have utility in the treatment of cancer. Therefore, various semi-synthetic derivatives of podophyllotoxin were produced and tested. However, up to now, the structural elucidation of new lignans still remains a huge task for experienced spectroscopists. Thus, the expert system SISTEMAT was developed in order to aid the specialists in this task and in structural elucidation of other chemical classes of natural compounds.

SISTEMAT[5,6] is an expert system developed for predicting the probable skeleton and substructures of natural products. It uses ^{1}H and ^{13}C nuclear magnetic resonance (^{1}H NMR and ^{13}C NMR) data and their bidimensional variants obtained from literature, in conjunction with data from source origin, to predict the chemical class and the type of carbonic skeleton of new compounds, which are fundamental data in the structural elucidation process. The SISTEMAT methodology also allows researchers to find out chemical information such as functional groups, mass and molecular formula, and

Received 11 September 2007; accepted 28 April 2008.

Address correspondence to Marcelo J. P. Ferreira, Instituto de Química, Universidade de São Paulo, Caixa Postal 26077, 05513-970, São Paulo, SP, Brazil. E-mail: marcelopena@mackenzie.br

oxidation number from the codified structures of compounds present in the database.

Various examples of expert systems in the structural determination area can be cited. Among them one can point out the DENDRAL,[7] ACCESS,[8] DARC/EPIOS,[9–13] SpecInfo,[14] and, more recently, the Assemble 2.0 and SENECA systems.[15–17] However, one of the greatest advantages of SISTEMAT in relation to the other systems is that structural codification system works very efficiently with data compression, thus enabling its use in microcomputers and its development specifically addressed to natural products chemistry.

The SISTEMAT database is being expanded by adding to it spectroscopic and source origin data of each new chemical class studied. Thus, through this system, several chemical classes were employed for structural elucidation like monoterpenes and iridoids,[18] lactonic and non-lactonic sesquiterpenes,[19] diterpenes,[20] and triterpenes.[21] These classes were also successfully analyzed using mainly data from ^{13}C NMR spectra and botanical origin.

As a part of our constant attempt toward natural product skeleton prediction through SISTEMAT, we have been worked with the lignan class for structural determination and skeleton elucidation. Therefore, the aim of this paper is to show the performance of the computer-aided structural elucidation of lignans and their skeleton prediction from the expert system SISTEMAT.

MATERIALS AND METHODS

The Lignan Database

The database built for this study contains 810 lignans with 490 ^{13}C NMR and 760 ^{1}H NMR spectra data collected from the literature (1975–2001). For the ^{13}C NMR data, chemical shifts and multiplicities were inserted in the database, however for the ^{1}H NMR data, only chemical shifts were stored. In both cases, the solvent used in the spectra acquisition was considered in the analysis. The lignans are distributed in 40 different skeletons, 1000 occurrences in 60 plant families, and were automatically coded by our DATASIS program.[22]

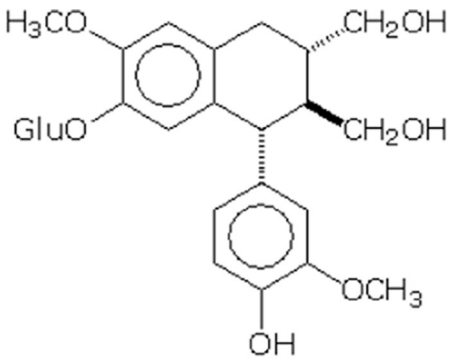
The SISTEMAT Programs

The programs MACRONO, SISCONST, C13MACH, H1MACH, and SISOCBOT were used for the lignan skeleton prediction. These programs propose, at the end of analysis, the probable skeleton compatible with the entered data. They were described in detail in previous papers,[18,23–25] however their main characteristics are described below.

Program MACRONO: This program is used for identification of common substituent groups such as acetate, angelate, tiglate, and so forth, bonded to the carbonic skeleton of natural products. From ^{13}C NMR data (chemical shifts and respective multiplicities) of the compound in question, the program searches in the database for the most probable substituent groups present in this compound and shows the user the selected substituent groups together with their average error range of chemical shifts. The average error range is the average deviation calculated from the chemical shifts of the sample in question and then matched with the chemical shifts values stored in the database. Therefore, the result will be more reliable as this range is minor. The MACRONO procedure is done because it removes all the ^{13}C NMR signals referring to the substituent groups. Thus, the chemical shift signals already removed from the substituent groups will not interfere with the skeleton analysis of the other programs of the SISTEMAT system, which carry out analysis and prediction of skeletons from ^{13}C NMR data. The database especially elaborated for this program contains 161 substituent groups more frequently encountered in natural product compounds.[24]

Program SISCONST: This program carries out ^{13}C NMR data analysis. From the chemical shifts and multiplicities of a compound, the program predicts its probable skeleton and identifies the substructures with their suitable attributed signals compatible with the experimental ^{13}C NMR data and shows them to the user. The program only selects substructures that have a minimum number of carbon atoms, demanded by the user, and their respective interlinked chemical shifts. This minimum number is suggested to deal with around half of the total carbon number of the chemical class in question. Hence, the minimum number for lignans is always nine carbon atoms.

The SISCONST program works matching the signals of the experimental spectrum with those of all spectra stored in the database. If a spectrum signal and its respective multiplicity are present in a determined carbon atom, the signals of the interlinked



Data from ^{13}C NMR spectra[27]: (C_5D_5N) 148.7 (s), 148.1 (s), 146.7 (s), 145.7 (s), 137.9 (s), 137.2 (s), 134.0 (s), 122.9 (d), 117.9 (d), 116.3 (d), 113.7 (d), 112.9 (d), 101.8 (d), 78.5 (d), 78.2 (d), 74.7 (d), 70.3 (d), 65.5 (t), 61.8 (t), 61.2 (t), 56.2 (q), 55.8 (q), 47.9 (d), 47.6 (d), 40.1 (d), 33.6 (t). Data from ^{1}H NMR spectra: (C_5D_5N): 7.09, 7.02, 7.01, 6.90, 6.86, 3.17, 3.17, 2.47, 2.47.

FIGURE 1 **Lignan used to exemplify the performance of the SISTEMAT program.**

carbons are matched with those from the experimental data. This searching process is repeated in order to obtain the largest substructures. All selected substructures will show a minimum number of the carbon atoms requested by the user as well as the chemical shifts compatible with the experimental ^{13}C NMR data. The skeleton probability of the substance is computed from the compounds that have a substructure containing at least half of the total carbon number. So, from each chosen compound and for its skeleton, the program calculates the skeleton probability as previously described.[26]

Programs C13MACH and H1MACH: These programs execute ^{13}C NMR and ^{1}H NMR data analysis, respectively. From the experimental data, chemical shifts and multiplicities from ^{13}C NMR and chemical shifts from ^{1}H NMR, the programs select *x*-compounds (*x* may change from 1 to 50) that display higher similarity index (SI) with the experimental data according to Bremser's system.[27] In this methodology, the chemical shifts do not need to be necessarily interlinked like those in the SISCONST program. Thus, from the selected compounds, the programs compute the skeleton probability.

The SISCONST, C13MACH, and H1MACH programs were able to differentiate among various stereoisomers furnishing correctly the relative configuration for many compounds. However, the system does not contain the ^{1}H NMR coupling constants data. Thus, the structures and substructures proposed by the system should be verified by the user through the analysis of the respective coupling constants.

Program SISOCBOT: This program was developed mostly to analyze source origin data based on chemosystematics. From the family and genus names of plants, marine organisms, and so forth, the program searches in the database for the number of occurrences of a specific skeleton exhibited in the family and genus of the species in question, and then, it predicts and shows the user the probability of finding a specific carbonic skeleton in the selected taxon.[25]

To exemplify the use of SISTEMAT programs, was selected a lignan (Fig. 1) isolated from *Cynomorium songaricum* (Cynomoriaceae), pertaining to aryltetralin skeleton.[28] The first step during the structural determination process, after the total insertion of ^{1}H and ^{13}C NMR chemical shifts and multiplicities of the lignan in question, is the use of the MACRONO

TABLE 1 **Skeleton Probability Showed by the SISTEMAT Programs and the Global Results**

Skeletons	SISCONST	C13MACH	H1MACH	Global results
02	84.1	62.8	52.23	55.3
03	4.1	—	—	—
19	3.3	—	—	—
01	—	18.5	35.6	18.8
11	—	18.5	11.9	10.0
15	—	18.5	—	—

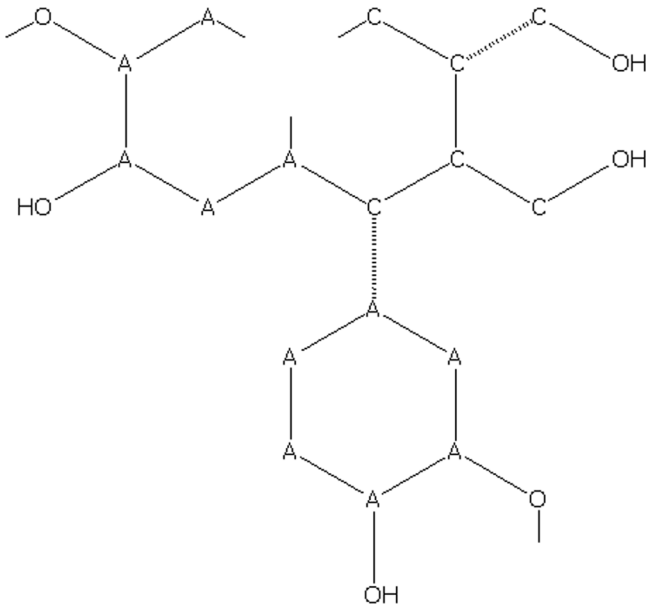


FIGURE 2 **Substructure with signal attribution exhibited by the program SISCONST. *The atoms designated by "A" represent the aromatic catbon atoms.**

program.[24] This program was selected as being capable of proposing the right substituent groups present. For example, in the lignan structure (Fig. 1) the groups selected were glucose (δ_C 101.8, 78.5, 78.2, 74.7, 70.3, 61.8) and two methoxyl (δ_C 56.2, 55.8) with a mean error of 0.73, 1.80, and 1.80, respectively.

Second, the program SISCONST[26] performs the ^{13}C NMR data analyses with the remaining chemical shifts and multiplicities that belong to the rest of the structure, predicting its probable skeleton and identifying the lignan substructures. In this test (Fig. 1), the program showed the skeleton probabilities (Table 1) and a substructure with ^{13}C NMR signal attributions (Fig. 2) at a chemical shift range of 2.0δ and gradient for substructure of 1.0δ.

The programs C13MACH and H1MACH[23] perform the ^{13}C and ^{1}H NMR data analyses, respectively. Both programs predict the probable skeleton of the test compound (Table 1). The programs also provide the compound structure with the highest SI in relation to the lignan tested. Figure 3 shows the structures I and II selected by programs C13MACH and H1MACH, respectively.

Finally, for the lignan test (Fig. 1) was used the program SISOCBOT,[25] which analyzes natural source data based on organism chemosystematics. However, the skeleton probability was not displayed because the family Cynomoriaceae is not inserted yet in our database. The sum of the results from these programs generates the global probability, which is computed from the weighted mean of the results of each program, at the end of all analyses. Thus, the following weights 0.9, 0.7, 0.9, and 0.3 were attributed for the programs SISCONST, C13MACH, H1MACH, and SISOCBOT, respectively. The weights attributed for each program vary from 0 to 1, and they were obtained through tests accomplished with a sampling of 30 compounds wherein was assessed the hit percent of each program. The global probability calculated for the lignan (Fig. 1) is shown in the global results column in Table 1.

RESULTS AND DISCUSSION

To test the performance and efficiency of the SISTEMAT programs, the data of 30 new lignans (Fig. 4) were collected randomly from literature.[29–51] It is important to notice that these lignans

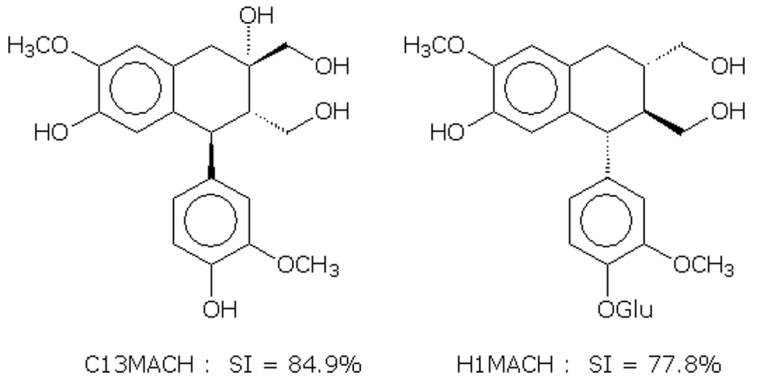


FIGURE 3 **Lignan structures proposed by C13MACH and HIMACH programs and their respectively similarly index (SI).**

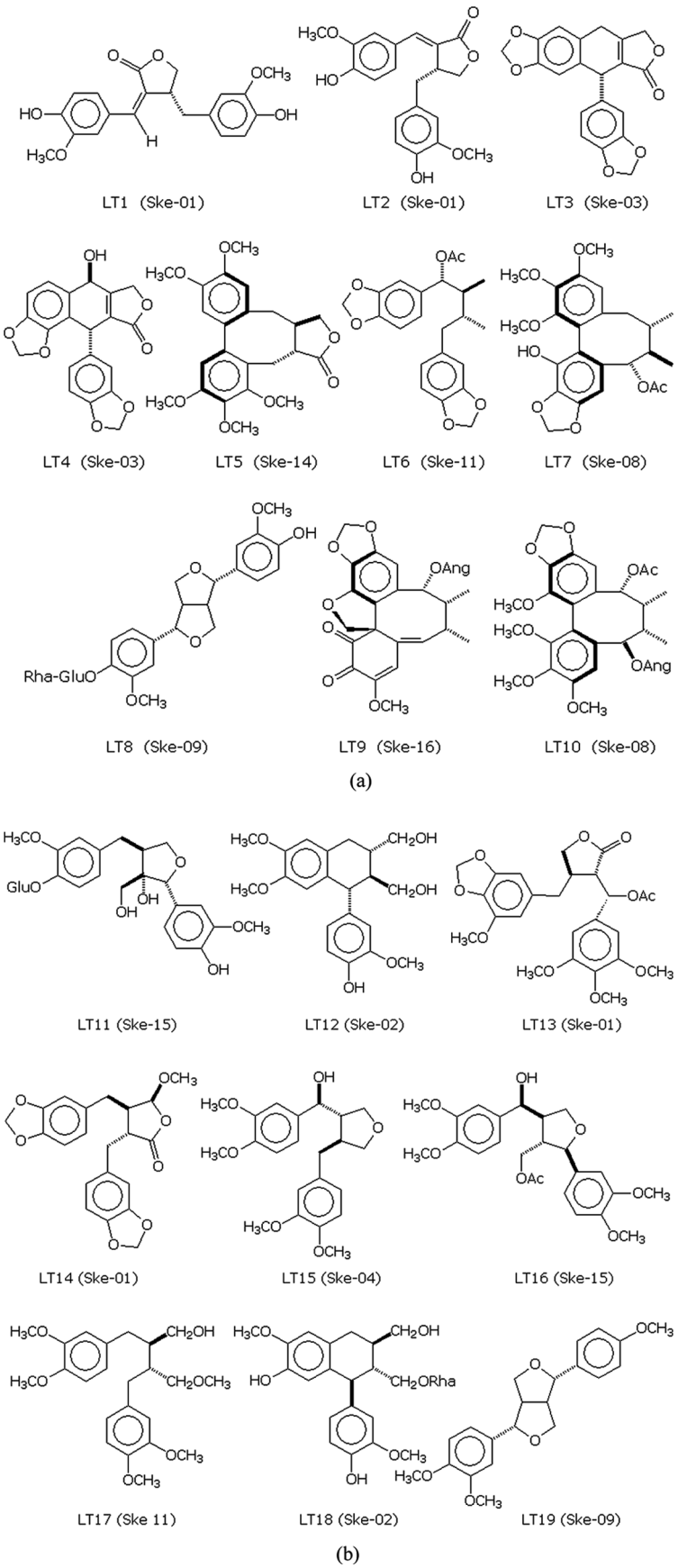
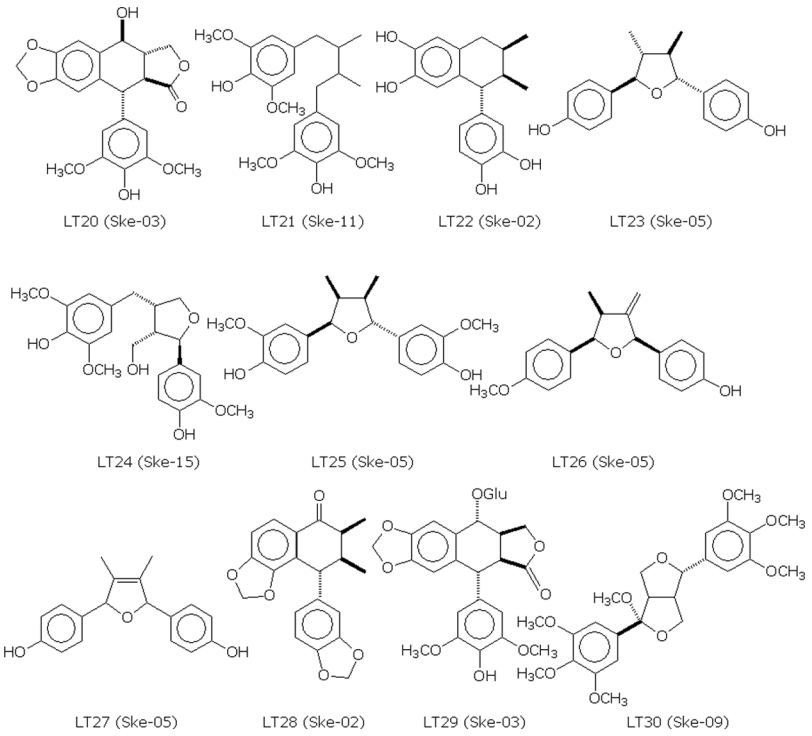


FIGURE 4 **Lignans (LT) used to test the SISTEMAT programs.**



(c)

FIGURE 4 **(Continued)**

were not stored in our database. Table 2 shows the ^{13}C NMR and ^{1}H NMR spectra data, the botanical data, and also exhibits the first three lignan skeletons (Fig. 5) proposed by the programs jointly with the global skeleton probability and respective references.

From the global results obtained and exhibited in Table 2, one can verify that the SISTEMAT programs showed the correct skeleton of the compounds in 80.0% of the cases (LT1-2, LT6-13, LT17–30).

By analyzing each program separately, the program MACRONO was applied for every lignan that exhibited substituent groups and was able to predict properly all of them. The program SISCONST showed the correct skeleton of the lignans in 66.7% of the cases (LT1-2, LT6-10, LT12, LT18–29) and incorrect skeleton of the compounds in 13.3% of the cases (LT3-5, LT15). In the remaining lignans tested (LT11, LT13–14, LT16–17, LT30), corresponding with 20.0% of the cases, the correct skeleton appeared at least in the first three skeletons analyzed by SISCONST, however this result may not be considered correct for skeleton predictions.

Although the program SISCONST has generated some wrong skeleton probabilities, it showed correct substructures for LT3, LT4, LT17, and LT30. SISCONST did not present correct substructures for LT5 because the number of lignans representing the same skeleton in the database is not significant (the skeleton 14 only has two compounds in the database). The program also did not provide correct substructures for LT15 (the skeleton 04 has 25 compounds in the database) and LT13 (the skeleton 03 has 58 compounds in the database) probably due to the similarity between the skeletons 04 and 15, and the skeletons 01 and 03, respectively. Finally, SISCONST failed to predict correct substructure for LT11 and LT16 (the skeleton 15 has 58 compounds in the database) and LT14 (the skeleton 01 has 108 compounds in the database) due to the substitution pattern in these cases being very related to the structures pertaining to the proposed skeletons. Therefore, for the least cases, the skeleton probability showed was wrong, too, because this probability is computed from the compounds that have a substructure containing at least half of the total carbon number, as previously described.[18]

TABLE 2 **Results Obtained from Each Program of the System SISTEMAT**

Compound	^{13}C NMR chemical shift, multiplicity, and botanical data	^{1}H NMR data	MACRONO	SISCONST	C13MACH	H1MACH	SISOCBOT	Global results	Refs.
LT1 Ske 01	($CDCl_3$): 174.6 s, 147.5 s, 146.5 s, 146.0 s, 144.5 s, 141.1 d, 129.6 s, 126.6 d, 126.4 s, 124.1 s, 121.8 d, 114.5 d, 113.8 d, 113.1d, 111.8 d, 70.0 t, 56.0 q, 55.9 q, 44.6 d, 40.9 t Umbelliferae- *Bupleurum salicifolium*	($CDCl_3$): 8.23, 7.02, 6.87 (X2), 6.71, 6.65, 6.55, 4.34, 4.13, 3.30, 2.86 (X2)	OCH_3: 56.0, 55.9	*Ske 01–90.1* Ske 02–3.5 Ske 03–1.5	*Ske 01–72.0* Ske 04–19.0 Ske 05–9.1	*Ske 01–100.0*	*Ske 01–83.3* Ske 06, Ske 07–8.3 each	*Ske 01–87.7* Ske 04–4.7 Ske 05–2.3	29
LT2 Ske 01	($CDCl_3$): 176.0 s, 147.7 s, 146.8 (X2) s, 144.8 s, 137.4 d, 129.8 s, 126.8 s, 125.9 d, 125.9 s, 124.0 d, 115.0 d, 114.8 d, 112.8 d, 111.7 d, 69.7 t, 56.1 q, 56.0 q, 39.8 d, 37.6 t Umbelliferae- *Bupleurum salicifolium*	($CDCl_3$): 7.52, 7.21, 7.03, 7.00, 6.87, 6.70, 6.64, 5.92, 5.53, 4.28 (X2), 3.84, 3.06, 2.63	OCH_3: 56.1, 56.0	*Ske 01–96.2* Other- 2.4 Ske 08–1.4	*Ske 01–61.1* Ske 02–29.5 Ske 09–9.4	*Ske 01–99.9* Ske 04–0.1	*Ske 01–83.3* Ske 06, Ske 07–8.3 each	*Ske 01–87.3* Ske 02–7.4 Ske 09–2.3	29
LT3 Ske 03	(DMSO-d_6): 171.9 s, 160.5 s, 147.3 s, 146.3 s, 146.2 s, 145.8 s, 137.7 s, 130.1 s, 125.3 s, 124.0 s, 121.0 d, 108.8 d, 108.3 d, 108.0 (X2) d, 100.8 t, 100.1 t, 71.2 t, 41.2 d, 28.5 t Acanthaceae- *Justicia neesii*	(DMSO-d_6): 6.84, 6.77, 6.68 (X2), 6.67, 5.97, 5.93, 5.92, 5.91, 5.03, 4.95, 4.72, 3.95, 3.68	—	Ske 06–40.1 Ske 02–20.2 Ske 10–12.4	Ske 01–42.1 Ske 04–38.7 Ske 11–19.2	Ske 09–99.9 *Ske 03–0.1*	Ske 11, Ske 06–33.3 each Ske 12–16.7 Ske 09–11.1	Ske 09–33.3 Ske 06–16.5 Ske 01–10.9	30
LT4 Ske 03	(DMSO-d_6): 171.6 s, 163.9 s, 147.2 s, 146.1 (X2) s, 144.5 s, 134.1 s, 131.9 s, 125.0 s, 121.2 d, 120.5 d, 118.9 s, 108.3 d, 108.1 d, 107.6 d, 101.2 t, 100.9 t, 70.4 t, 62.6 d, 36.9 d Acanthaceae- *Justicia neesii*	(DMSO-d_6): 7.20, 6.94, 6.76, 6.64, 6.62, 5.97, 5.94, 5.93 (X2), 5.67, 5.15, 5.01, 4.79	—	Ske 06–70.7 Ske 02–13.5 Ske 10–5.5	Ske 01–50.0 Ske 04, Ske 11–20.3 each Ske 15–9.7	Ske 09–99.4 Ske 01, Ske 11, *Ske 03*, Ske 13–0.2 each	Ske 11, Ske 06–33.3 each Ske 12–16.7 Ske 09–11.1	Ske 09–33.1 Ske 06–26.3 Ske 01–12.6	30
LT5 Ske 14	($CDCl_3$): 176.0 s, 151.8 s, 150.7 s, 149.2 s, 147.7 s, 141.9 s, 136.3 s, 132.8 s, 131.2 s, 126.8 s, 114.5 d, 112.7 d, 110.1 d, 70.0 t, 61.1 q, 60.8 q, 56.3 q, 49.8 d, 47.0 d, 34.4 t, 24.3 t Apiaceae- *Steganotaenia araliacea*	($CDCl_3$): 6.72, 6.69, 6.52, 4.38, 3.70 3.68, 2.69, 2.15 (X4)	OCH_3: 61.1, 60.8, 56.3	Ske 02–53.6 Ske 03–23.1 Ske 19, Ske 06–4.7 each	Ske 09–40.1 Ske 02–29.7 Ske 01–19.6	Ske 08–100.0	*Ske 14–66.6* Ske 01–33.3	Ske 08–32.1 Ske 02–24.6 Ske 09–10.1	31

(*Continued*)

TABLE 2 Continued

Compound	^{13}C NMR chemical shift, multiplicity, and botanical data	^{1}H NMR data	MACRONO	SISCONST	C13MACH	H1MACH	SISOCBOT	Global results	Refs.
LT6 Ske 11	($CDCl_3$): 170.2 s, 147.8 s, 147.6 s, 147.3 s, 145.7 s, 134.8 s, 133.8 s, 121.8 d, 121.2 d, 109.2 d, 108.0 d (X2), 107.6 d, 101.0 t, 100.7 t, 79.2 d, 41.5 t, 40.0 d, 34.8 d, 21.1 q, 14.0 q, 9.8 q Magnoliaceae- *Talauma ovata*	($CDCl_3$): 6.54 (X6), 5.94 (X2), 5.90 (X2), 5.52, 2.40, 1.57 (X2), 0.91 (X3), 0.78 (X3)	OAc: 170.2 s, 21.1 q	*Ske 11–84.1* Ske 03–7.0 Ske 08–6.2	Ske 04, *Ske 11*–40.3 each Ske 01–19.4	*Ske 11–97.1* Ske 04–2.1 Ske 15, Ske 26, Ske 01–0.2 each	Ske 05–100.0	*Ske 11–68.3* Ske 05, Ske 04–10.7 each Ske 01–4.9	32
LT7 Ske 08	($CDCl_3$): 170.0 s, 150.3 s, 148.9 s, 146.5 s, 141.2 s, 136.1 s, 135.6 s, 133.7 s, 133.3 s, 119.0 s, 117.0 s, 107.1 d, 102.7 d, 101.2 t, 82.6 d, 60.8 q, 59.8 q, 55.8 q, 41.5 d, 38.6 t, 35.0 d, 21.6 q, 19.7 q, 14.8 q Schisandraceae- *Kadsura matsudai*	($CDCl_3$): 6.46, 6.37, 5.95, 5.92, 5.50, 2.61 (X2), 2.07 (X2), 1.06 (X3), 0.88 (X3)	OCH_3: 60.8, 59.8, 55.8 OAc: 170.0 s, 21.6 q	*Ske 08–72.7* Ske 25–8.5 Ske 03–6.4	Ske 25–30.8 *Ske 08–22.5* Ske 15–19.2	*Ske 08–100.0*	*Ske 08–56.7* Ske 16–40.0 Ske 17–3.3	*Ske 08–67.3* Ske 25–10.4 Ske 15–4.8	33
LT8 Ske 09	(DMSO-d_6): 149.4 s, 147.7 s, 146.0 s, 145.5 s, 135.5 s, 132.5 s, 118.8 d, 117.9 d, 115.3 d, 115.2 d, 110.5 d, 110.0 d, 100.1 d, 98.2 d, 85.4 d, 85.2 d, 77.9 d, 76.9 d, 75.8 d, 72.4 d, 71.2 t, 71.1 t, 70.7 d, 70.5 d, 69.9 d, 68.4 d, 60.6 t, 55.7 q, 55.5 q, 53.9 d, 53.8 d, 18.1 q Malvaceae- *Hibiscus syriacus*	(DMSO-d_6): 7.02, 6.92, 6.87, 6.85, 6.74, 6.71, 5.22, 4.99, 4.68, 4.62, 4.13 (X2), 3.92, 3.75 (X2), 3.68, 3.63, 3.54, 3.44, 3.43, 3.37, 3.29, 3.19, 3.18, 3.04 (X2), 1.11 (X3)	OCH_3: 55.7 Glu: 100.1 d, 77.9 d, 76.9 d, 75.8 d, 70.7 d, 60.6 t Rha: 98.2 d, 75.8 d, 72.4 d, 70.7 d, 70.5 d, 18.1 q	Ske 09–74.8 Ske 05–18.7 Ske 15–2.7	*Ske 09–90.9* Ske 15–9.1	*Ske 09–100.0*	*	*Ske 09–78.9* Ske 05–7.2 Ske 15–5.3	34
LT9 Ske 16	($CDCl_3$): 189.2 s, 175.8 s, 167.1 s, 151.0 s, 150.6 s, 142.8 s, 140.7 d, 137.5 d, 132.9 s, 130.6 s, 129.5 s, 127.7 s, 126.1 d, 118.8 s, 102.3 t, 101.9 d, 79.9 t, 79.2 d, 66.7 s, 55.7 q, 43.9 d, 30.5 d, 20.3 q, 19.5 q, 15.6 q, 11.0 q Schisandraceae- *Kadsura interior*	($CDCl_3$): 6.67, 6.50, 6.02 (X2), 5.83, 5.82, 4.84, 4.40, 3.03, 1.88, 1.04 (X3), 1.02 (X3)	OCH_3: 55.7 OAng: 167.1, 137.5, 127.7, 20.3, 15.6	*Ske 16–80.9* Ske 08–12.4 Ske 02–3.7	*Ske 16–71.6* Ske 06–19.1 Ske 08–9.3	*Ske 16–59.6* Ske 08–40.4	Ske 08–53.3 *Ske 16–40.0* Ske 08, Ske 17–3.3 each	*Ske 16–67.3* Ske 08–25.4 Ske 06–4.8	35

LT10 Ske 08	($CDCl_3$): 170.0 s, 166.7 s, 151.8 s, 151.5 s, 148.6 s, 141.7 s, 141.2 s, 138.9 d, 135.9 s, 133.1 s, 131.3 s, 127.8 s, 121.1 (X2) s, 110.4 d, 102.2 d, 101.0 t, 80.7 (X2) d, 60.6 q, 60.2 q, 59.3 q, 56.0 q, 38.7 (X2) d, 20.7 (X2) q, 19.9 q, 15.6 (X2) q Schisandraceae- *Kadsura interior*	($CDCl_3$): 6.71, 6.44, 5.95 (X2), 5.84, 5.73, 2.22, 2.12, 1.03 (X3), 0.94 (X3)	OCH_3: 59.3, 56.0, 60.2, 60.6; OAc: 170.0 s, 20.7 q; OAng: 166.7 s; 138.9 d, 127.8 s, 20.7 q, 15.6 q	*Ske 08–85.0* Ske 06–6.9 Ske 10–4.8	*Ske 08–56.4* Ske 11–35.2 Ske 03–8.4	*Ske 08–79.7* Ske 18–11.6 Ske 01, Ske 13, Ske 07–2.9 each	Ske 08–53.3 *Ske 16–40.0* Ske 08, Ske 17–3.3 each	*Ske 08–72.8* Ske 11–8.8 Ske 16–4.3	35
LT11 Ske 15	(CD_3OD): 150.9 s, 149.1 s, 148.7 s, 146.4 s, 136.9 s, 129.6 s, 122.3 d, 121.6 d, 118.4 d, 116.0 d, 114.5 d, 112.7 d, 103.1 d, 85.7 d, 83.2 s, 78.2 d, 77.8 d, 75.0 d, 71.8 t, 71.4 d, 64.5 t, 62.5 t 56.8 q, 56.3 q, 51.7 d, 35.1 t Valerianaceae- *Valeriana officinalis*	(CD_3OD): 7.09, 6.92, 6.90, 6.77, 6.73 (X2), 4.82, 4.05, 3.79, 3.63, 3.60, 3.12, 2.59, 2.54	OCH_3: 56.3, 56.8 Glu: 103.1 d, 78.2 d, 77.8 d, 75.0 d, 71.4 d, 62.5 t	Ske 02–72.3 *Ske 15–13.8* Ske 19–5.1	*Ske 15–41.9* Ske 11–29.4 Ske 02–10.0	*Ske 15–100.0*	*	*Ske 15–47.1* Ske 02–25.7 Ske 11–7.4	36
LT12 Ske 02	(CD_3OD) 148.4 s, 148.3 s, 148.2 s, 146.0 s, 138.5 s, 134.0 s, 130.0 s, 123.2 d, 116.0 d, 114.6 d, 113.8 d, 112.8 d, 65.9 t, 62.1 t, 56.5 q, 56.4 (X 2) q, 48.0 d, 46.0 d, 40.0 d, 33.5 t Sterculiaceae- *Scaphopetalum thonneri*	(CD_3OD): 6.76, 6.71, 6.69, 6.63, 6.28, 3.88, 3.71, 3.68, 3.42, 2.81, 2.80, 2.01, 1.78	OCH_3: 56.4, 56.5	*Ske 02–80.1* Ske 03–7.8 Ske 19–4.6	*Ske 02–81.1* Ske 15–18.9	*Ske 02–77.4* Ske 11–22.5	*	*Ske 02*–70.9 Ske 11–7.8 Ske 08–6.1	37
LT13 Ske 01	($CDCl_3$): 175.9 s, 169.2 s, 159.4 s, 153.4 (X2) s, 149.2 s, 134.3 s, 132.9 s, 131.2 s, 108.6 s, 108.5 d, 102.4 (X2) d, 102.2 d, 101.6 t, 73.7 d, 72.0 t, 60.8 d, 56.7 q, 56.2 q, 50.8 (X2) q, 39.5 t, 37.6 d, 21.0 q Hernandiaceae- *Hernandia sonora*	($CDCl_3$): 6.39 (X2), 6.14, 6.04, 5.97, 5.94, 5.91, 4.31, 3.96, 2.78, 2.75, 2.46, 2.29	OCH_3: 56.7 OAc: 21.0 q, 169.2 s	Ske 03–42.7 *Ske 01–18.4* Ske 02–12.4	Ske 09–50.7 Ske 04–20.1 *Ske 01–19.6*	*Ske 01–94.2* Ske 08–5.2 Ske 16–0.6	Ske 15–66.7 *Ske 01–33.3*	*Ske 01–44.7* Ske 03–16.1 Ske 09–12.7	38
LT14 Ske 01	($CDCl_3$): 177.7 s, 147.7 (X2) s, 146.3 (X2) s, 131.6 s, 130.8 s, 122.0 d, 121.8 d, 110.0 t, 109.1 d, 108.9 d, 108.4 d, 108.0 (X2) d,	($CDCl_3$): 6.57, 6.55, 6.39, 6.38, 6.29, 6.24, 5.90 (X2), 5.80 (X2), 5.00, 2.95, 2.63, 2.62, 2.41, 2.26, 2.25	OCH_3: 57.0	Ske 11–40.3 *Ske 01–19.1* Ske 04–17.5	*Ske 01–39.7* Ske 11–29.5 Ske 04–20.7	Ske 11–77.2 *Ske 01–20.5* Ske 04–1.8	Ske 05–66.7 Ske 02–33.3	Ske 11–45.1 *Ske 01–22.7* Ske 04–11.4	39

(*Continued*)

TABLE 2 Continued

Compound	^{13}C NMR chemical shift, multiplicity, and botanical data	^{1}H NMR data	MACRONO	SISCONST	C13MACH	H1MACH	SISOCBOT	Global results	Refs.
	100.9 t, 57.0 q, 46.8 d, 46.4 d, 37.5 t, 36.7 t Aristolochiaceae- *Aristolochia elegans*								
LT15 Ske 04	($CDCl_3$): 152.0 s, 149.0 s, 148.5 s, 147.4 s, 135.5 s, 133.0 s, 120.5 d, 118.0 d, 112.0 d, 111.4 d, 111.1 d, 109.1 d, 82.8 d, 73.0 t, 61.0 t, 55.9 (X4) q, 52.6 d, 42.4 d, 33.3 t Aristolochiaceae- *Aristolochia elegans*	($CDCl_3$): 6.89, 6.85, 6.82, 6.80, 6.74, 6.71, 4.82, 4.06, 3.93, 3.78, 3.76, 2.93, 2.75, 2.58, 2.43	OCH_3: 55.9	Ske 15–66.2 Ske 02–23.9 Ske 03–3.0	Ske 15–55.7 Ske 01, Ske 11–9.1 each	Ske 15–100.0	Ske 05–66.7 Ske 02–33.3	Ske 15–67.4 Ske 02–13.5 Ske 05–7.1	39
LT16 Ske 15	($CDCl_3$): 170.7 s, 149.2 s, 149.1 s, 148.8 s, 148.6 s, 135.1 s, 133.6 s, 118.9 d, 118.6 d, 110.9 d (X2), 109.2 d (X2), 83.9 d, 75.8 d, 70.1 t, 63.8 t, 55.9 (X2) q, 49.9 d, 49.0 d, 20.7 q Magnoliaceae- *Magnolia praecocissima*	($CDCl_3$): 6.92, 6.88, 6.86, 6.84, 6.81 (X2), 4.59, 4.52, 4.34, 3.98, 3.85 (X2), 2.50, 2.10	OCH_3: 55.9 OAc: 20.7q, 170.7s	Ske 02–44.4 *Ske 15–39.5* Ske 03–8.0	*Ske 15–41.6* Ske 05–30.2 Ske 09–19.2	Ske 09–76.5 Ske 05–20.2 *Ske 15–3.3*	Ske 09–34.5 Ske 25–27.6 *Ske 15–24.1*	Ske 09–34.8 *Ske 15–26.7* Ske 02, Ske 05–14.6 each	40
LT17 Ske 11	($CDCl_3$): 148.9 (X2) s, 147.4 (X2) s, 133.4 s, 133.2 s, 121.1 (X2) d, 112.4 (X2) d, 111.3 (X2) d, 71.1 t, 60.9 t, 58.9 q, 55.9 (X2) q, 55.8 (X2) q, 44.6 d, 42.5 d, 36.2 t, 35.8 t Euphorbiaceae- *Phyllanthus niruri*	($CDCl_3$): 6.65 (X6), 3.46 (X4), 2.60 (X4), 1.90 (X2)	OCH_3: 58.9, 55.9, 55.8	Ske 01–44.2 *Ske 11–35.7* Ske 02–10.2	*Ske 11–61.5* Ske 01–19.0 Ske 15–10.2	*Ske 11–87.1* Ske 02–12.4 Ske 04–0.5	Ske 10–57.1 Ske 06–28.6 Ske 05–14.3	*Ske* 11–54.8 Ske 01–21.2 Ske 02–7.3	41
LT18 Ske 02	(CD_3OD): 149.2 s, 147.2 s, 146.1 s, 145.2 s, 138.1 s, 133.9 s, 128.9 s, 123.2 d, 117.1 d, 116.1 d, 113.4 d, 112.4 d, 102.3 d, 73.8 d, 72.5 d, 72.3 d, 70.1 d, 67.9 t, 65.3 t, 56.3 q, 48.3 d, 45.5 d, 40.0 d, 33.6 t, 17.9 q Polygonaceae- *Polygonum aviculare*	(CD_3OD): 6.75, 6.66, 6.63, 6.59, 6.16, 3.85, 3.81, 3.71, 3.63, 3.34, 3.10, 2.83 (X2), 2.02, 1.86	OCH_3: 56.3 Rha: 102.3 d, 73.8 d, 72.5 d, 72.3 d, 70.1 d, 17.9 q	*Ske 02–77.2* Ske 19–6.5 Ske 03–5.4	*Ske 02–65.4* Ske 01–17.8 Ske 04, Ske 15–8.4 each	*Ske 02–77.8* Ske 11–22.2	*	*Ske 02*–66.1 Ske 11–7.5 Ske 01–4.9	42

LT19 Ske 09	($CDCl_3$): 159.1 s, 149.1 s, 148.5 s, 133.5 s, 133.0 s, 127.2 (X2) d, 118.2 d, 113.8 (X2) d, 110.9 d, 109.1 d, 85.7 d, 85.5 d, 71.7 t, 71.4 t, 55.8 q, 55.1 (X2) q, 54.1 d, 54.0 d Annonaceae- *Rollinia membranacea*	($CDCl_3$): 7.10 (X7), 4.73 (X2), 4.22 (X2), 3.90 (X2), 3.05 (X2)	OCH_3: 55.8, 55.1	*Ske 09–81.2* Ske 05–15.8 Ske 04, Ske 15–1.4 each	*Ske 09–100.0*	*Ske 09–100.0*	Ske 20–100.0	*Ske* 09–83.2 Ske 20–10.8 Ske 05–5.1	43
LT20 Ske 03	(DMSO-d_6): 177.2 s, 147.7 (X2) s, 145.6 (X2) s, 134.5 s, 134 s, 133.1 s, 131.0 s, 107.3 d, 106.7 (X2) d, 104.4 d, 100.3 t, 68.9 t, 67.1 d, 56.0 q, 44.2 d, 43.1 d, 42.6 d Berberidaceae- *Podophyllum versipellis*	(DMSO-d_6): 7.15, 6.63 (X2), 6.17, 5.92 (X2), 4.60, 4.48, 4.44, 3.91, 3.38, 2.65	OCH_3: 56.0	*Ske 03–64.0* Ske 02–24.8 Ske 10–3.2	Ske 09–29.1 *Ske 03–20.7* Ske 02–20.0	*Ske 03–100.0*	*Ske 03–100.0*	*Ske* 03– 68.6 Ske 02–13.0 Ske 09–7.3	44
LT21 Ske 11	($CDCl_3$): 146.8 (X4) s, 132.9 (X2) s, 132.7 (X2) s, 105.6 (X4) d, 56.3 q, 39.4 (X2) t, 39.0 (X2) d, 16.3 (X2) q Zygophyllaceae- *Porlieria chilensis*	($CDCl_3$): 6.32 (X4), 2.70 (X2), 2.27 (X2), 1.75 (X2), 0.83 (X6)	OCH_3: 56.3	*Ske 11–87.8* Ske 02–7.5 Ske 03–2.6	*Ske 11–71.9* Ske 02–10.1 Ske 05–9.0	Ske 09–55.3 *Ske 11–36.9* Ske 08–7.8	*	*Ske 11*–58.0 Ske 09–17.8 Ske 03–11.5	45
LT22 Ske 02	((CD_3)$_2$ CO): 145.2 s, 144.0 s, 143.7 (X2) s, 140.1 s, 130.3 s, 128.0 s, 121.2 d, 117.6 d, 116.8 d, 115.7 d, 115.5 d, 50.8 d, 41.5 d, 35.4 t, 29.8 d, 16.2 q, 16.0 q Zygophyllaceae- *Larrea tridentata*	((CD_3)$_2$ CO): 6.71, 6.56, 6.44, 6.43, 6.26, 3.51, 2.80, 2.37, 2.02, 1.88, 0.87 (X6)	—	*Ske 02–94.9* Ske 19–2.9 Ske 22–1.0	Ske 15–49.6 *Ske 02–22.0* Ske 01, Ske 11, Ske 09–9.5 each	*Ske 02–99.8* Ske 08–0.2	*	*Ske 02–68.1* Ske 03–10.7 Ske 15–9.9	46
LT23 Ske 05	((CD_3)$_2$ CO): 157.6 (X2) s, 134.0 (X2) s, 128.3 (X4) d, 115.8 (X4) d, 88.7 (X2) d, 51.9 (X2) d, 13.9 (X2) q Zygophyllaceae- *Larrea tridentata*	((CD_3)$_2$ CO): 7.24 (X4), 6.82 (X4), 4.59 (X2), 1.74 (X2), 1.0 (X6)	—	*Ske 05–87.5* Ske 09–12.5	*Ske 05–100.0*	*Ske 05–92.7* Ske 11–7.2 Ske 15–0.1	*	*Ske 05–82.9* Ske 03–10.7 Ske 09–4.0	47
LT24 Ske 15	($CDCl_3$): 147.1 (X2) s, 146.7 s, 145.1 s, 134.7 s, 133.0 s, 131.5 s, 118.8 d, 114.2 d, 108.3 d, 105.3 (X2) d, 82.7 d, 72.8 t, 60.9 t, 56.3 q, 55.9 q, 52.7 d, 42.5 d, 33.8 t Acanthaceae- *Justicia glauca*	($CDCl_3$): 6.87 (X2), 6.80, 6.42 (X2), 4.77, 4.04, 3.82 (X2), 3.77, 2.93, 2.73, 2.53, 2.42	OCH_3: 56.3, 55.9	*Ske 15–56.5* Ske 02–23.7 Ske 03–4.3	*Ske 15–61.2* Ske 04–18.8 Ske 09–10.7	Ske 02–54.8 *Ske 15–45.2*	Ske 11, Ske 06–33.3 each Ske 12–16.7 Ske 09–11.1	*Ske 15–48.0* Ske 02–25.2 Ske 11–5.9	48

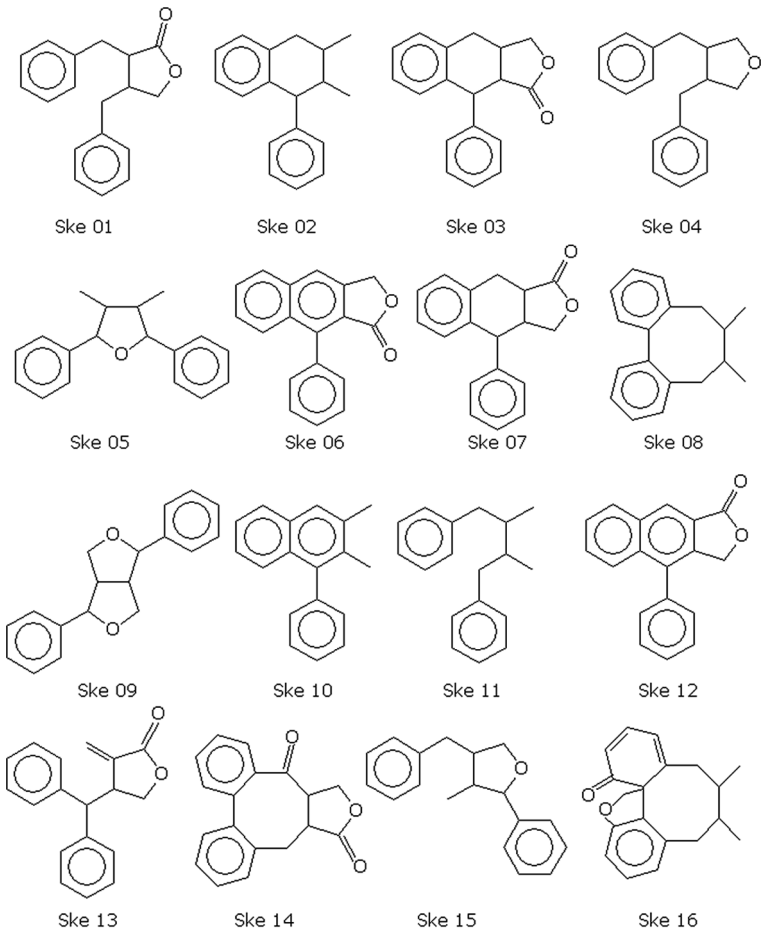
(*Continued*)

TABLE 2 Continued

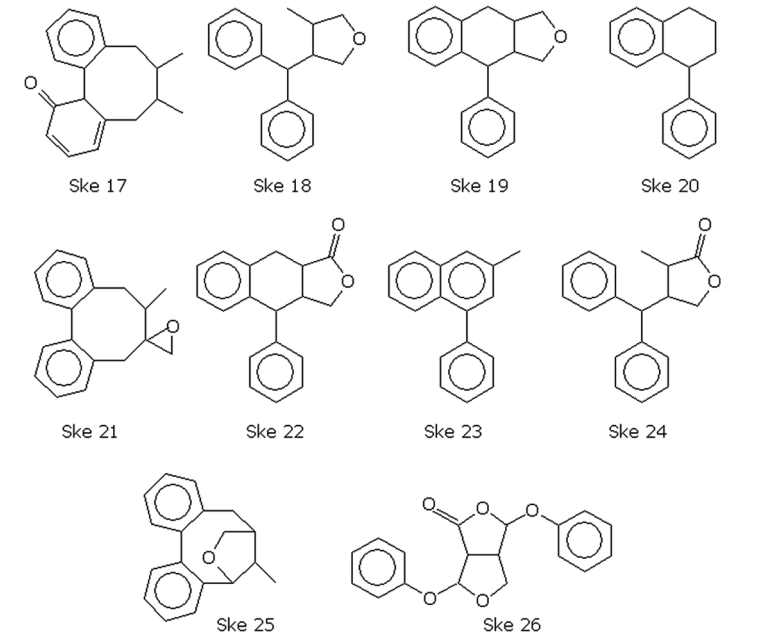
Compound	^{13}C NMR chemical shift, multiplicity, and botanical data	^{1}H NMR data	MACRONO	SISCONST	C13MACH	H1MACH	SISOCBOT	Global results	Refs.
LT25 Ske 05	(C_6 D_6): 147.1 s, 146.9 s, 145.9 s, 145.2 s, 135.5 s, 133.0 s, 119.6 d, 119.1 d, 114.7 d, 114.5 d, 109.1 d, 108.9 d, 85.9 d, 84.7 d, 47.9 d, 43.6 d, 11.9 q, 9.5 q (methoxy data- not given by authors) Zygophyllaceae- *Larrea tridentata*	(C_6 D_6): 7.12, 7.07, 7.06, 6.92, 6.82, 6.78, 5.37, 4.65, 2.25, 2.10, 0.86 (X3), 0.61 (X3)	—	*Ske 05–70.8* Ske 02–21.7 Ske 04–3.1	Ske 09–30.4 *Ske 05–21.4* Ske 04–19.9	*Ske 05–92.6* Ske 11–7.3 Ske 08–0.1	*	*Ske 05–57.9* Ske 09–11.2 Ske 02–6.9	47
LT26 Ske 05	((CD_3)$_2$ CO): 159.9 s, 159.7 s, 158.0 s, 133.1 s, 132.5 s, 129.8 (X2) d, 128.3 (X2) d, 115.9 (X2) d, 114.1 (X2) d, 106.8 t, 84.0 d, 82.6 d, 43.7 d, 17.5 q (methoxy data- not given by authors) Zygophyllaceae- *Larrea tridentata*	((CD_3)$_2$ CO): 7.32 (X2), 7.31 (X2), 6.93 (X2), 6.87 (X2), 5.25, 5.09, 5.02, 4.64, 3.10, 0.73 (X3)	—	*Ske 05–100.0*	Ske 09–29.7 *Ske 05–29.4* Ske 11–20.1	*Ske 05–56.7* Ske 09–42.5 Ske 04–0.6	*	*Ske 05–57.7* Ske 09–22.3 Ske 11–8.6	47
LT27 Ske 05	((CD_3)$_2$ CO): 157.8 (X2) s, 134.4 (X2) s, 131.7 (X2) s, 129.1 (X4) d, 115.8 (X4) d, 91.8 (X2) d, 10.4 (X2) q Zygophyllaceae- *Larrea tridentata*	((CD_3)$_2$CO): 7.15 (X4), 6.82 (X4), 5.64 (X2), 1.49 (X6)	—	*Ske 05–77.5* Ske 23–22.5	*Ske 05–46.8* Ske 11–18.4 Ske 09–17.9	*Ske 05–70.9* Ske 01, Ske 09–14.2 each Ske 04–0.6	*	*Ske 05–59.4* Ske 09–10.2 Ske 23–9.4	47
LT28 Ske 02	($CDCl_3$): 198.7 s, 151.7 s, 147.8 s, 146.3 s, 145.7 s,	($CDCl_3$): 7.73, 6.83, 6.70, 6.57, 6.48, 5.93 (X2), 5.92	—	*Ske 02–75.2* Ske 03–14.8	*Ske 02–60.1* Ske 01–19.9	*Ske 02–100.0*	Ske 01–50.0 Ske 03–30.8	*Ske 02–72.2* Ske 01–12.9	49

	136.0 s, 127.5 s, 123.7 s, 122.7 d, 121.0 d, 108.5 d, 108.2 d, 107.6 d, 102.0 t, 101.0 t, 45.7 d, 43.3 d, 40.9 d, 15.4 q, 12.6 q Cupressaceae-*Chamaecyparis obtusa*	(X2), 4.14, 2.81, 2.31, 1.08 (X3), 1.02 (X3)		Ske 06–3.1	Ske 24–10.2		*Ske 02–7.7* **	Ske 03–8.0	
LT29 Ske 03	(CD_3OD): 177.1 s, 149.0 (X2) s, 148.6 (X2) s, 135.9 s, 133.5 s, 132.7 s, 132.5 s, 110.2 d, 109.7 (X2) d, 109.6 d, 102.6 t, 80.4 d, 72.9 t, 56.9 (X2) q, 46.5 d, 45.1 d, 40.6 d Berberidaceae-*Sinopodophyllum amodi*	(CD_3OD): 7.37, 6.46, 6.40 (X2), 6.0, 5.91, 5.06, 4.69, 4.56, 4.21, 2.98, 2.96	OCH_3: 56.9	*Ske 03–56.7* Ske 02–31.3 Ske 06–3.6	Ske 01–39.9 Ske 09–28.9 Ske 05–20.7	*Ske 03–98.8* Ske 22–1.2	Ske 03–75.0 Ske 09, Ske 05, Ske 06–8.3 each **	*Ske 03–58.0* Ske 02, Ske 01–10.0 each Ske 09–8.1	50
LT30 Ske 09	($CDCl_3$): 153.4 (X2) s, 153.1 (X2) s, 137.6 s, 137.3 s, 136.6 s, 133.2 s, 110.3 s, 103.9 (X2) d, 103.0 (X2) d, 88.0 d, 70.6 t, 69.8 t, 61.1 q, 61.0 q, 57.0 d, 56.4 (X2) q, 56.3 (X2) q, 53.1 d, 49.0 q Asteraceae- *Artemisia caruifolia*	($CDCl_3$): 6.71 (X2), 6.56 (X2), 4.48, 4.12, 4.07, 3.86, 3.31, 3.09, 3.04	OCH_3 : 61.1, 61.0, 56.4 (X2), 56.3 (X2), 49.0	Ske 02–29.4 *Ske 09–22.4* Ske 19–15.5	*Ske 09–42.4* Ske 05–29.8 Ske 15–18.6	*Ske 09–80.0* Ske 15–19.6 Ske 03–0.4	*Ske 09*, Ske 04–40.0 Ske 15–20.0	*Ske 09–47.8* Ske 15–13.5 Ske 02–9.4	51

Glu, glucose; **Rha**, rhamnose; **OAng**, angelate; **OAc**, acetate; **Xyl**, xylose.
*Family not inserted in SISTEMAT.
**Genus not found in SISTEMAT.



(a)



(b)

FIGURE 5 **Skeletons (Ske) proposed by the programs.**

The program C13MACH showed the correct skeleton of the compounds in 63.3% of the cases (LT1-2, LT6, LT8-12, LT14, LT16–19, LT21, LT23–24, LT27–28, LT30), a little less percent value obtained by the SISCONST (66.7%). Generally, the program C13MACH is not as accurate as the SISCONST[18,21] because the latter predicts the skeletons by selecting only lignans that contain a minimum number of carbon atoms interlinked, relating the chemical shifts and multiplicities of the carbons inserted in the database with each lignan's carbon tested. Hence the program also furnishes the structures or substructures corresponding with the correct skeleton. In contrast, the program C13MACH recognizes the skeletons, matching the chemical shifts and multiplicities of the lignan test against the stored spectral data in the database showing the most likely skeletons.

According to the Table 2, C13MACH exhibited incorrect skeleton of the lignans in 16.7% of the cases (LT3-5, LT15, LT29). In 20.0% of the cases (LT7, LT13, LT20, LT22, LT25–26), the correct skeletons were encountered at least in the first three skeletons proposed by C13MACH.

During these analyses, the program H1MACH was the most accurate one for lignan skeleton predictions, probably due to the greater number of ^{1}H NMR spectra data inserted in the database compared with ^{13}C NMR spectra data. This program showed the correct skeleton of the lignans in 73.3% of the cases (LT1-2, LT6-13, LT17–20, LT22–23, LT25–30). The lignans LT5 and LT15 were not correctly identified by H1MACH representing 6.7% of the wrong analyses. Moreover, the program was not able to predict correct skeletons for LT3-4, LT14, LT16, LT21, and LT24 as the most probable, but the correct skeleton was presented as one of the first three options in 20.0% of the cases.

On the contrary, the program SISOCBOT showed the correct skeleton of the lignans only in 40.0% of the cases (LT1-2, LT5, LT7, LT10, LT20, LT29–LT30). The search was only done at the genus level present in the database. The analysis carried out by this program was prejudiced because only in 66.7% of the cases could the program be evaluated, because the systematic data were not yet completed in the database. The incorrect skeleton was predicted in 60.0% of the cases where in 20.0% of them (LT9, LT13, LT16, and LT28) the correct skeleton was predicted among the three first skeletons proposed by the program.

CONCLUSIONS

The results of the analyses of lignans carried out in this work show a good performance of the programs composing the system SISTEMAT for identification and skeleton prediction with 80.0% hits. From the global results or individual results obtained with each program, it can be concluded that this system may be introduced as a valuable tool for the process of lignan skeleton determination, therefore increasing the spectral interpretation capacity of the researchers and, consequently, reducing the spent time for them during the structural determination process.

It is noteworthy that the program H1MACH will be introduced like a helpful tool to investigate possible skeletons of new lignans or other aromatic classes isolated once it exhibits better performance among the programs tested in this work. On the other hand, the analysis of aliphatic natural products, such as terpenoids, shows a high complexity when compared with that of the aromatic compounds, due to the high number of chemical shifts and the various proton couplings, which makes less accurate the program analysis.

From information such as substituent groups, skeleton type, substructures, and so forth, obtained by the SISTEMAT programs, the structural elucidation carried out by the structure generator should be processed more efficiently and quickly, once those data are used as a restriction module for the structure generator that is being developed for the expert system SISTEMAT. In the last months, the expert system SISTEMAT and the structure generator have been rewritten in JAVA interface and the system presentation will be carried out, in the future, at a specific website. The restriction module procedure was applied with success in the MOLGEN and LSD generators[52–54] and should be followed by our system avoiding the combinatorial explosion problem observed in other expert systems developed up to now.[7,11–12,55–56]

ACKNOWLEDGMENTS

This work was supported by grants from the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP). The authors thank Antônio J.C.

Brant for helpful discussion during the preparation of the manuscript. The authors M.J.P.F. and V.P.E. acknowledge the CNP_9 scholarship.

REFERENCES

1. Moss, G. P. Nomenclature of lignans and neolignans. *Pure Appl. Chem.* **2000**, *72*(8), 1493–1523.
2. Ward, R. S. Lignans, neolignans, and related compounds. *Nat. Prod. Rep.* **1995**, *12*(2), 183–205.
3. Ward, R. S. Lignans, neolignans, and related compounds. *Nat. Prod. Rep.* **1997**, *14*(1), 43–74.
4. Gottlieb, O. R. *New Natural Products and Plant Drugs with Pharmacological, Biological or Therapeutical Activity*; Springer-Verlag: Berlin, 1987; 227–248.
5. Gastmans, J. P.; Furlan, M.; Lopes, M. N.; Borges, J. H. G. Emerenciano, V. P. A inteligência artificial aplicada à química de produtos naturais. O programa SISTEMAT. Parte I-Bases teóricas. *Quim. Nova* **1990**, *13*(1), 10–16.
6. Gastmans, J. P.; Furlan, M.; Lopes, M. N.; Borges, J. H. G. Emerenciano, V. P. A inteligência artificial aplicada à química de produtos naturais. O programa SISTEMAT. Parte II-Organização do programa e aplicativos. *Quim. Nova* **1990**, *13*(1), 75–80.
7. Lindsay, R. K.; Buchanan, B. G.; Feigenbaum, E. A.; Lederberg, J. *Applications of Artificial Intelligence for Organic Chemistry: The DENDRAL Project*; McGraw-Hill: New York, 1980.
8. Bremser, W. Multidimensional spectroscopy. *Magn. Reson. Chem.* **1985**, *23*(12), 1056–1071.
9. Carabedian, M.; Dagane, J.; Dubois, J.E. Elucidation by progressive intersection of ordered substructures from carbon-13 nuclear magnetic resonance. *Anal. Chem.* **1988**, *60*(20), 2186–2192.
10. Dubois, J. E.; Carabedian, M. Single-resonance subspectra/substructure investigations of the carbon-13 DARC databank. Representation of local and global topological knowledge. *J. Chem. Inf. Comput. Sci.* **1991**, *31*(4), 557–564.
11. Carabedian, M.; Dubois, J. E. A combined model of multi-resonance subspectra substructure and DARC topological – structure representation – local and global knowledge in the ^{13}C NMR DARC database. *J. Chem. Inf. Comput. Sci.* **1991**, *31*(4), 564–574.
12. Carabedian, M.; Dubois, J. E. Inferring extended virtual knowledge from an EPIOS conversion graph of overlapping substructures. *J. Chem. Inf. Comput. Sci.* **1994**, *34*(4), 701–706.
13. Carabedian, M.; Dubois, J.E. Large virtual enhancement of a C-13 NMR database. A structural crowning extrapolation method enabling spectral data transfer. *J. Chem. Inf. Comput. Sci.* **1998**, *38*(2), 100–107.
14. Will, M.; Fachinger, W.; Richert, J. R. Fully automated structure elucidation – A spectroscopist's dream comes true. *J. Chem. Inf. Comput. Sci.* **1996**, *36*(2), 221–227.
15. Munk, M.E. Computer-based structure determination: Then and now. *J. Chem. Inf. Comput. Sci.* **1998**, *38*(6), 997–1009.
16. Badertscher, M.; Korytko, A.; Schulz, K. P.; Madison, M.; Munk, M. E.; Portmann, P.; Junghans, M.; Fontana, P.; Pretsch, E. Assemble 2.0: A structure generator. *Chemom. Intell. Lab. Syst.* **2000**, *51*(1), 73–79.
17. Steinbeck, C. SENECA: A platform-independent, distributed, and parallel system for computer-assisted structure elucidation in organic chemistry. *J. Chem. Inf. Comput. Sci.* **2001**, *41*(6), 1500–1507.
18. Ferreira, M. J. P.; Rodrigues, G. V.; Emerenciano, V. P. MONOREG – an expert system for structural elucidation of monoterpenes. *Can. J. Chem.* **2001**, *79*(12), 1915–1925.
19. Oliveira, F. C.; Ferreira, M. J. P.; Núñez, C. V.; Rodriguez, G. V.; Emerenciano, V. P. ^{13}C NMR spectroscopy of eudesmane sesquiterpenes. *Prog. Nucl. Magn. Reson. Spectrosc.* **2000**, *37*(1–2), 1–45.
20. Alvarenga, S. A. V.; Gastmans, J. P.; Rodrigues, G. V.; Emerenciano, V. P. Ditregra-an auxiliary program for structural determination of diterpenes. *Spectroscopy* **1997**, *13*(3), 227–249.
21. Macari, P. A. T.; Gastmans, J. P.; Rodrigues, G. V.; Emerenciano, V. P. An expert system for structure elucidation of triterpenes. *Spectroscopy* **1994/1995**, *12*(3), 139–166.
22. Emerenciano, V. P.; Rodrigues, G. V.; Gastmans, J. P. Um método de codificação semi-automático assistido por computador para o SISTEMAT. *Quim. Nova* **1993**, *16*(4), 431–434.
23. Costantin, M. B.; Ferreira, M. J. P.; Rodrigues, G. V.; Emerenciano, V. P. Use of an expert system in lignan skeleton prediction from ^{1}H NMR data. *Internet Electron. J. Mol. Des.* **2004**, *3*(12), 759–770.
24. Ferreira, M. J. P.; Oliveira, F. C.; Alvarenga, S. A. V.; Macari, P. A. T.; Rodrigues, G. V.; Emerenciano, V. P. Automatic identification by ^{13}C NMR of substituent groups bonded in natural product skeletons. *Comput. Chem. (Oxford)* **2002**, *26*(6), 601–632.
25. Ferreira, M. J. P.; Alvarenga, S. A. V.; Macari, P. A. T.; Rodrigues, G. V.; Emerenciano, V. P. A program for terpenoid skeleton prediction based on botanical information. *Biochem. Syst. Ecol.* **2003**, *31*(1), 25–43.
26. Fromanteau, D. L. G.; Gastmans, J. P.; Vestri, S. A.; Emerenciano, V. P.; Borges, J. H. G. A constraints generator in structural determination by microcomputer. *Comput. Chem. (Oxford)* **1993**, *17*(4), 369–378.
27. Bremser, W.; Klier, M.; Meyer, E. Mutual assignment of subspectra and substructures – A way to structure elucidation by ^{13}C NMR spectroscopy. *Org. Magn. Reson.* **1975**, *7*(2), 97–106.
28. Jiang, Z. H.; Tanaka, T.; Sakamoto, M.; Jiang, T.; Kuono, I. Studies on a medicinal parasitic plant: Lignans from the stems of *Cynomorium songaricum*. *Chem. Pharm. Bull.* **2001**, *49*(8), 1036–1038.
29. Braun, A. E.; Reyes, R. E.; González, A. G. Structural elucidation and conformational analysis of new lignan butenolides from the leaves of *Bupleurum salicifolium*. *Tetrahedron* **1994**, *50*(17), 5203–5210.
30. Rajasekhar, D.; Subbaraju, G. V.; Ravicumar, K.; Chandramohan, K. *Justicia* lignans V. Three new *β*-apolignans from *Justicia neesii* Ramamoorthy. *Tetrahedron* **1998**, *54*(43), 13227–13236.
31. Hicks, R. P.; Sneden, A. T. Neoisostegane, a new bisbenzocyclooctadiene lignan lactone from *Steganotaenia araliaceae* Hochst. *Tetrahedron Lett.* **1983**, *24*(29), 2987–2990.
32. Stefanello, M. E. A.; Alvarenga, M. A.; Toma, I. N. New neolignans from *Talauma ovata*. *Fitoterapia* **2002**, *73*(2), 135–139.
33. Kuo, Y. H.; Li, S. Y; Wu, M. D.; Huang, R. L.; Kuo, L. M. Y., Chen, C. F. A new anti-HBeAg lignan, Kadsumarin A, from *Kadsura matsudai* and *Schizandra arisanensis*. *Chem. Pharm. Bull.* **1999**, *47*(8), 1047–1048.
34. Lee, S-J.; Yun, Y-S.; Lee, I-K.; Ryoo, I-J; Yun, B-S.; Yoo, I-D. An antioxidant lignan and other constituents from the root bark of *Hibiscus syriacus*. *Planta Med.* **1999**, *65*(7), 658–660.
35. Chen, D. F.; Zhang, S. X.; Kozuka, M.; Sun, Q. Z.; Feng, J.; Wang, Q.; Mukainaka, T.; Nobukuni, Y.; Tokuda, H.; Nishino, H.; Wang, H. K.; Natschke, S. L. M.; Lee, K. H. 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*(9), 1242–1245.
36. Schumacher, B.; Scholle, S.; Hölzl, J.; Khudeir, N.; Hess, S.; Müller, C. E. Lignans isolated from Valerian: Identification and characterization of a new olivil derivative with partial agonistic activity at A1 adenosine receptors. *J. Nat. Prod.* **2002**, *65*(10), 1479–1485.
37. Vardamides, J. C.; Azabaze, A. G. B.; Nkengfack, A. E.; van Heerden, F. R. Fomum, Z. T.; Ngando, T. M.; Conrad, J.; Vogler, B.; Kraus, W. Scaphopetalone and scaphopetalumate, a lignan and a triterpene ester from *Scaphopetalum thonneri*. *Phytochemistry* **2003**, *62*(4), 647–650.
38. Udino, L.; Abaul, J.; Bourgeois, P.; Gorrichon, L.; Duran, H.; Zedde, C. Lignans from the seeds of *Hernandia sonora*. *Planta Med.* **1999**, *65*(3), 279–281.
39. Wu, T. S.; Tsai, Y. L.; Damu, A. G.; Kuo, P. C.; Wu, P. L. Constituents from the root and stem of *Aristolochia elegans*. *J. Nat. Prod.* **2002**, *65*(11), 1522–1525.

40. Takahashi, H.; Yoshioka, S.; Kawano, S.; Azuma, H.; Fukuyama, Y. Lignans and sesquiterpenes from *Magnolia praecocissima*. *Chem. Pharm. Bull.* **2002**, *50*(4), 541–543.
41. Satyanarayana, P.; Subrahmanyam, P.; Viswanatham, K. N. New seco- and hydroxy-lignans from *Phyllanthus niruri*. *J. Nat. Prod.* **1988**, *51*(1), 44–49.
42. Kim, H. J.; Woo, E. R.; Park, H. A novel lignan and flavonoids from *Polygonum aviculare*. *J. Nat. Prod.* **1994**, *57*(5), 581–586.
43. Sahpaz, S.; Villaescusa, L.; Hocquemiller, R.; Cavé, A. Rioclarine et Membranacine, deux nouvelles acetogenins bis-tetrahydrofuraniques des graines de Rollinia membranacea. *J. Nat. Prod.* **1993**, *56*(3), 351–356.
44. Yu, P. Z.; Wang, L. P.; Chen, Z. N. A new podophyllotoxin-type lignan from *Dysosma versipellis* var. *tomentosa*. *J. Nat. Prod.* **1991**, *54*(5), 1422–1424.
45. Torres, R.; Urzua, A.; Modak, B. Isopregomisin, a 1,4-bis-(phenyl)-2,3-dimethylbutane lignan from *Porlieria chilensis*. *J. Nat. Prod.* **1989**, *52*(2), 402–403.
46. Konno, C.; Xue, H. Z.; Lu, Z. Z.; Ma, B. X.; Erdelmeier, C. A. J.; Che, C. T.; Cordell, G. A.; Soejarto, D. D.; Waller, D. P.; Fong, H. H. S. 1-Aryl tetralin lignans from *Larrea tridentata*. *J. Nat. Prod.* **1989**, *52*(5), 1113–1117.
47. Konno, C.; Lu, Z. Z.; Xue, H. Z.; Ma, B. X.; Erdelmeier, C. A. J.; Meksuriyen, D.; Che, C. T.; Cordell, G. A.; Soejarto, D. D.; Waller, D. P.; Fong, H. H. S. Furanoid lignans from *Larrea tridentata*. *J. Nat. Prod.* **1990**, *53*(2), 396–406.
48. Subbaraju, G. V.; Kumar, K. K. K.; Raju, B. L.; Pillai, K. R. Justiciresinol, a new furanoid lignan from Justicia glauca. *J. Nat. Prod.* **1991**, *54*(6), 1639–1641.
49. Kuo, Y. H.; Chen, C. H.; Chiang, Y. M. Three novel and one new lignan, chamaecypanones A, B, obtulignolide and isootobanone from the heartwood of *Chamaecyparis obtusa* var. *formosana*. *Tetrahedron Lett.* **2001**, *42*(38), 6731–6735.
50. Zhao, C.; Huang, J.; Nagatsu, A.; Ogihara, Y. Two new podophyllotoxin glucosides from *Sinopodophyllum emodi* (WALL.) YING. *Chem. Pharm. Bull.* **2001**, *49*(6), 773–775.
51. Ma, C.; Nakamura, N.; Min, B. S.; Hattori, M. Triterpenes and lignans from *Artemisia caruifolia* and their cytotoxic effects on Meth-A and LLC tumor cell lines. *Chem. Pharm. Bull.* **2001**, *49*(2), 1 83–187.
52. Benecke, C.; Grund, R.; Hohberger, R.; Kerber, A.; Laue, R.; Wieland, T. MOLGEN^{+}, a generator of connectivity isomers and stereoisomers for molecular structure elucidation. *Anal. Chim. Acta* **1995**, *314*(3), 141–147.
53. Universität Bayreuth. MOLGEN, V 3.1. Universität Bayreuth, Bayreuth, Austria.
54. Nuzillard, J.-M.; Emerenciano, V. P. Automatic structure elucidation through data base search and 2D NMR spectral analysis. *Nat. Prod. Commun.* 2006, *1*(1), 57–64.
55. Carhart, R. E.; Smith, D. H.; Gray, N. A. B.; Nourse, J. G.; Djerassi, C. GENOA: A computer program for structure elucidation utilizing overlapping and alternative substructures. *J. Org. Chem.* **1981**, *46*(8), 1708–1718.
56. Jaspars, M. Computer assisted structure elucidation of natural products using two-dimensional NMR spectroscopy. *Nat. Prod. Rep.* **1999**, *16*(2), 241–247.