Internet Electronic Journal of Molecular Design **2003**, *2*, 000–000
BioChem Press http://www.biochempress.com

^{13}C NMR Pattern Recognition of Guaiane Sesquiterpenes

Marcelo J. P. Ferreira,[1] Francimeiry C. Oliveira,[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 xxx; Preprint published xxx; Accepted xxx ; Published xxx

Internet Electron. J. Mol. Des. **2003**, *2*, 000–000

Abstract

Motivation. Sesquiterpenes are a class of naturally occurring substances that show various biological and pharmacological properties such as inhibitors of cell growth, antifungal, antibacterial, antimalarial, antiviral, etc activities. However, jointly with many biological properties, this class exhibits a greater diversity and complexity of structures, becoming a great challenge in the structural elucidation of its compounds for spectroscopists. Thus, the development of an expert system containing a set of rules for structural pattern recognition, based on ^{13}C NMR spectral data, for new compounds is of utmost importance for the natural product researchers.

Method. SISTEMAT is an expert system developed to aid in the structural determination of natural products. In this study we have investigated the application of system for the ^{13}C NMR pattern recognition of guaiane sesquiterpenes. By using a database containing 200 substances, and the programs of the system SISTEMAT, various ^{13}C NMR heuristic rules for structural elucidation were obtained. These rules were evaluated with a set of 25 new guaianes recently published in literature.

Results. The prediction performance of the system from the tests executed with the compounds shows that the system was able to propose substructures in 96.0% of the studied cases, where in 98.1% of these are overlapping substructures.

Conclusions. SISTEMAT is a powerful tool for structural elucidation with many potential applications in natural products' field. In the present study we have demonstrated the predictive ability and applicability of a ^{13}C NMR pattern recognition method for the guaiane sesquiterpenes.

Availability. The software used in this study can be consulted by a contact with the corresponding author.

Keywords. ^{13}C NMR; guaiane sesquiterpenes; pattern recognition; substructure identification.

Abbreviations and notations

^{13}C NMR, Carbon-13 nuclear magnetic resonance	EP, epoxide
EN, double bond	OXY, ether ciclic
OXO, carbonyl group	

1 INTRODUCTION

It is well-established that ^{1}H and ^{13}C NMR are the most powerful analytical methods for solving structure-elucidation problems. These techniques are relevant when one treats structure elucidation of new natural products, due to the great diversity and the structural complexity found within these classes of substances [1-2].

Structure determination of terpenoids attracts widespread interest because they represent an important group of naturally occurring substances, and many of them exhibit pharmacological

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

Copyright © 2003 BioChem Press

properties [3-5]. Among the terpenoids, sesquiterpenes are a class providing constant challenges for structure elucidation, due to the countless biogenetic pathways that can produce several types of carbon skeletons [6]. Concerning this point, we decide to create, through the expert system SISTEMAT [7,8], a specialist module for sesquiterpenoids, aimed at facilitating the process of structure elucidation.

As there are hundreds of substances of this class with reported ^{13}C NMR data, we began building this specialist module using one of the most representative skeletons of the sesquiterpenes, the guaiane type (Figure 1).

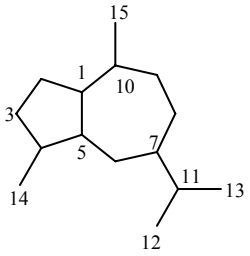

Figure 1. Guaiane skeleton

The aim of this paper is to show how SISTEMAT can be used to obtain useful rules for ^{13}C spectral analysis (^{13}C NMR pattern recognition) and can be used as an auxiliary tool in the process of structure elucidation for guaianes.

2 MATERIALS AND METHODS

2.1 The Expert System SISTEMAT

The expert system denominated SISTEMAT has been developed by our group since the 90s [7,8]. The major goal of this system is to become an auxiliary tool for chemists of natural products, thus enabling these researchers to achieve the most likely carbon skeletons of the compounds more quickly and effectively. In our approach was used the division of compounds into skeletal types because it is one of the fundamental points in the process of structural determination of occurring naturally substances. Knowledge about carbon skeletons helps to reduce the search time and avoids a combinatorial explosion in the generator of structures [9]. Thus, several chemical classes of natural products have already been studied using SISTEMAT, for example, monoterpenes, eudesmane sesquiterpenes, diterpenes, triterpenes and flavonoids [10-15].

The specialist module in sesquiterpenes with a guaiane skeleton was built by coding ^{13}C NMR data collected from the literature. This procedure resulted in 200 substances that were inserted into the SISTEMAT. The utilized codes contain information on chemical formula, molecular mass, and types of carbon atoms at a specific position that can be later accessed through a special command of the system.

2.2 The ^{13}C NMR Pattern Recognition

In the system SISTEMAT, the search for heuristic rules for pattern recognition is done through the programs TIPCARB [13,16] and PICKUP [16]. The TIPCARB program can determine the carbon atoms which are present in each position of a skeleton. This information helps in the search for heuristic rules, once these define whether or not the skeleton is substituted, and the kind of its substituents. This could also be done manually by a careful analysis of the literature, but the huge data volume makes this task unfeasible for obtaining heuristic rules.

After the position of each carbon atom and the types of substituents having been defined, the fragments, denominated substructures, are coded in the PICKUP program. PICKUP acts as a heuristic function and searches for the spectral patterns of each skeletal type. The approach used in producing the set of structural building units is a three-step process. The first step consists of identifying certain structural features to allow the characterization of a given skeleton or substructure. The second step consists of choosing the substructures and atomic groupings from the menu in order to estimate the ^{13}C NMR shift ranges for a particular group of substances. The PICKUP module then performs the search of the database for the chemical shift range from ^{13}C NMR data of the carbons in the substructure.

After chemical shift estimation, the obtained information is evaluated in relation to its degree of recognition with the complete database allowing one to affirm that a certain group of chemical shifts characterizes a certain probability of the occurrence of a substructure in the compound. In summary, the TIPCARB program indicates the substructure that should be selected, and the PICKUP program obtains the chemical shift ranges of the carbon atoms of that substructure, besides exhibiting the degree of recognition of the chemical shifts within the database.

3 RESULTS AND DISCUSSION

The PICKUP program afforded various chemical shift ranges that characterize several substructures present in guaianes. These results and the percentual of recognition are presented in Table 1. The use of these groups of chemical shifts can be applied in the processes of structure elucidation for new guaianes. Thus, in order to check if the chemical shift ranges obtained were useful for structure determination of new substances, 25 guaianes (Fig. 2) were randomly selected from the literature. These compounds were deliberately inserted in the database for calculating the ranges shown in Table 1 and were also used to test the efficiency of obtaining substructures from the chemical shifts of new substances. The substances had their ^{13}C NMR data submitted to the program, that proposed corresponding substructures by means of research through characteristic chemical shift ranges. The results of these tests are shown in Table 2, where only the first fifteen chemical shifts of each compound are given, after the chemical shifts of the substituents having been previously identified and removed by the program MACRONO [17].

Table 1. ^{13}C NMR shifts ranges for several substructures

Substructure	N^{o} C	^{13}C NMR shift ranges		% Recognition
[6,10OXY]	1	54.0 – 53.3	d	100.0
	5	67.9 – 67.3	d	
	6	75.9 – 75.2	d	
	10	76.4 – 73.4	s	
[6βOH; 7α,10α-OXY]	6	74.3 – 72.0	d	100.0
	7	87.2 – 86.0	s	
	8	29.1 – 28.3	t	
	9	32.1 – 31.4	t	
	10	84.9 – 82.8	s	
[6β,7βEP]	6	65.0 – 59.5	d	100.0
	7	70.4 – 65.8	s	
	8	21.7 – 20.0	t	
	9	29.3 – 28.6	t	
	10	34.4 – 32.5	d	
[10,11OXY]	7	44.0 – 31.0	d	100.0
	10	76.4 – 73.3	s	
	11	76.8 – 74.5	s	
	12	31.1 – 23.0	q	
[2,8OR; 10,11OXY]	2	77.0 – 73.6	d	100.0
	7	44.0 – 39.9	d	
	8	71.8 – 64.3	d	
	10	76.4 – 73.4	s	
	11	76.8 – 74.5	s	
[5,11OXY]	5	92.5 – 90.4	s	100.0
	7	46.0 – 45.4	d	
	11	81.8 – 80.5	s	
[4βOH; 14α]	1	51.0 – 45.4	d	93.7
	2	26.5 – 21.0	t	
	3	41.1 – 33.0	t	
	4	80.5 – 69.3	s	
	14	25.3 – 12.5	q	
[11OH]	7	53.6 – 45.5	d	100.0
	8	28.0 – 21.6	t	
	9	37.2 – 27.4	t	
	10	44.8 – 34.0	d	
	11	74.3 – 72.9	s	

Table 1. Continued

Substructure	N° C	^{13}C NMR shift ranges		% Recognition
[6OH; 7,10Peroxy]	6	75.7 – 73.5	d	100.0
	7	84.0 – 82.6	s	
	8	21.7 – 20.0	t	
	9	34.0 – 30.8	t	
	10	77.5 – 77.0	s	
[6,8OH; 7,10Peroxy]	6	71.0 – 67.5	d	100.0
	7	88.0 – 86.6	s	
	8	66.9 – 66.0	d	
	9	40.7 – 39.6	t	
	10	79.1 – 78.5	s	
[1,3,5,9EN]	1	145.3 – 144.1	s	100.0
	4	135.1 – 134.5	s	
	5	136.0 – 132.8	s	
	10	134.6 – 133.1	s	
[1(5)EN]	1	146.8 – 138.8	s	100.0
	2	36.2 – 34.2	t	
	4	46.5 – 43.4	d	
	5	141.1 – 136.0	s	
	10	36.7 – 33.4	d	
[1(5)EN; 2OXO]	1	145.3 – 140.2	s	100.0
	2	210.8 – 204.0	s	
	3	46.0 – 42.5	t	
	4	37.9 – 32.0	d	
	5	181.7 – 167.0	s	
[1(5)EN;6OH; 7,10Peroxy]	1	143.8 – 140.7	s	100.0
	5	181.7 – 176.8	s	
	6	75.5 – 67.5	d	
	7	88.1 – 82.6	s	
	10	79.0 – 77.0	s	
[1(5),6EN]	1	142.4 – 142.0	s	100.0
	5	136.0 – 135.4	s	
	6	117.5 – 117.1	d	
	10	150.8 – 150.2	s	
[5,8OXY; 8OH]	1	52.7 – 51.7	d	100.0
	5	86.9 – 84.6	s	
	8	103.8 – 97.8	s	

Table 1. Continued

Substructure	N° C	^{13}C NMR shift ranges			% Recognition
[1(5),6EN; 2OXO]	1	142.5	140.2	s	100.0
	2	210.7	209.6	s	
	5	170.1	167.0	s	
	6	118.0	113.1	d	
	7	168.3	164.2	s	
[1(10)EN]	1	173.6 –	139.8	s	100.0
	4	39.5 –	36.9	d	
	5	49.0 –	45.0	d	
	6	34.9 –	28.0	t	
	10	135.8 –	121.6	s	
[1(10)EN; 11OH]	1	148.3 –	139.8	s	100.0
	10	128.8 –	121.6	s	
	11	75.5 –	73.1	s	
[1(10),11EN]	1	173.6 –	142.0	s	100.0
	4	39.2 –	36.2	d	
	5	49.0 –	46.2	d	
	7	51.2 –	42.7	d	
	10	135.8 –	125.5	s	
	11	152.1 –	148.8	s	
[1(10),2,4,5,8EN]	2	135.8 –	125.5	d	100.0
	5	147.1 –	138.5	s	
	7	149.3 –	135.6	s	
	9	115.1 –	114.1	d	
	10	139.1 –	135.5	s	
[1(10)EN; 14OXO; 14OR]	1	170.1 –	166.3	s	100.0
	5	49.0 –	48.5	d	
	10	126.0 –	125.5	s	
	14	173.8 –	169.1	s	
[1(10),3EN; 2OXO]	1	136.0 –	132.8	s	100.0
	2	197.0 –	196.1	s	
	3	133.6 –	132.6	d	
	4	175.5 –	172.0	s	
	10	171.0 –	146.1	s	
[5,6EP]	5	71.8 –	70.8	s	100.0
	6	60.5 –	59.7	d	

Table 1. Continued

Substructure	N° C	^{13}C NMR shift ranges		% Recognition
[2OXO; 3,5EN]	2	205.5 – 205.1	s	100.0
	3	130.8 – 130.0	d	
	4	174.8 – 169.0	s	
	5	143.3 – 143.3	s	
	6	126.6 – 126.3	d	
[3OXO; 4EN]	1	46.0 – 42.7	d	100.0
	2	43.0 – 40.0	t	
	3	208.0 – 210.9	s	
	4	141.4 – 137.5	s	
	5	176.7 – 167.5	s	
[1βOH; 3OXO; 4EN]	1	80.0 – 78.3	s	97.0
	2	51.2 – 49.9	t	
	3	207.0 – 204.8	s	
	4	137.0 – 136.1	s	
	5	175.3 – 174.0	s	
[3OXO; 4,11EN]	3	209.6 – 200.5	s	100.0
	4	147.0 – 137.6	s	
	5	176.6 – 171.1	s	
	11	150.3 – 134.8	s	
	12	125.0 – 109.0	t	
[6EN; 10αOH]	6	121.7 – 116.0	d	100.0
	7	166.3 – 149.5	s	
	8	28.0 – 25.0	t	
	9	42.8 – 37.0	t	
	10	77.5 – 71.6	s	
[6EN; 8,10OH]	6	117.5 – 113.0	d	100.0
	7	168.0 – 166.0	s	
	8	69.3 – 68.7	d	
	9	54.5 – 45.6	t	
	10	71.9 – 70.7	s	
[7(11)EN; 8OXO]	7	136.6 – 133.5	s	100.0
	8	209.0 – 194.6	s	
	11	155.3 – 133.8	s	
	12	22.6 – 20.6	q	
	13	23.0 – 22.0	q	

Table 1. Continued

Substructure	N° C	^{13}C NMR shift ranges		% Recognition
[4OR; 10(14)EN]	1	54.0 – 42.7	d	100.0
	4	81.5 – 80.3	s	
	5	59.5 – 47.2	d	
	10	152.5 – 149.1	s	
	14	111.0 – 108.0	t	
[4,6OH; 10(14)EN]	4	80.8 – 80.3	s	100.0
	6	72.5 – 71.6	d	
	10	152.5 – 149.1	s	
	14	111.0 – 108.0	t	
[10(14)EN; 11OH]	7	49.3 – 45.5	d	100.0
	8	30.5 – 26.0	t	
	9	41.3 – 34.6	t	
	10	153.7 – 152.5	s	
	11	74.5 – 72.3	s	
	14	109.9 – 106.2	t	
[6,8CYCLO]	6	15.0 – 12.0	d	100.0
	7	25.0 – 21.0	d	
	8	13.0 – 10.0	d	
[5,6; 11,12EP]	11	60.7 – 60.4	s	100.0
	12	54.9 – 54.7	t	

The ^{13}C NMR data shown in Table 2 were recorded in $CDCl_3$, except tests XXII and XXIII which were recorded in C_6D_6.

The system used here was able to propose substructures in 96.0% of the studied cases. The unique negative result (guaiane XX) was due to the inexistence of precise and characteristic rules for the compound. A more detailed analysis of the cases, where the program was able to propose by overlapping substructures, gave a 98.1% success rate, and, in 52.0% of the cases, it was possible to build up a complex molecular structure.

An imprecise assignment occurred in test IX, where three substructures were proposed with incorrect positioning of the hydroxyl group, but, jointly with the correct substructures, the result always displayed one among the three as the correct one. In many cases the overlapping of two or more proposed substructures resulted in the correct assignment of the substance structure, for example, in tests I, III, IV, VI and XVIII.

It is also necessary to comment that in tests XI–XVII and XIX, the program was able to correctly supply the complete structure of the compound, however it competes with the analyst in considering the correct stereochemistry of the hydroxyl groups. This problem is due to the absence of the ^{13}C

NMR pattern recognition that characterizes the correct configuration of these groups. In test XXV the substituent group acetate was indicated correctly by the program MACRONO. In the previous paper [17], this program was widely evaluated in the identification of substituent groups bonded in diverse natural products.

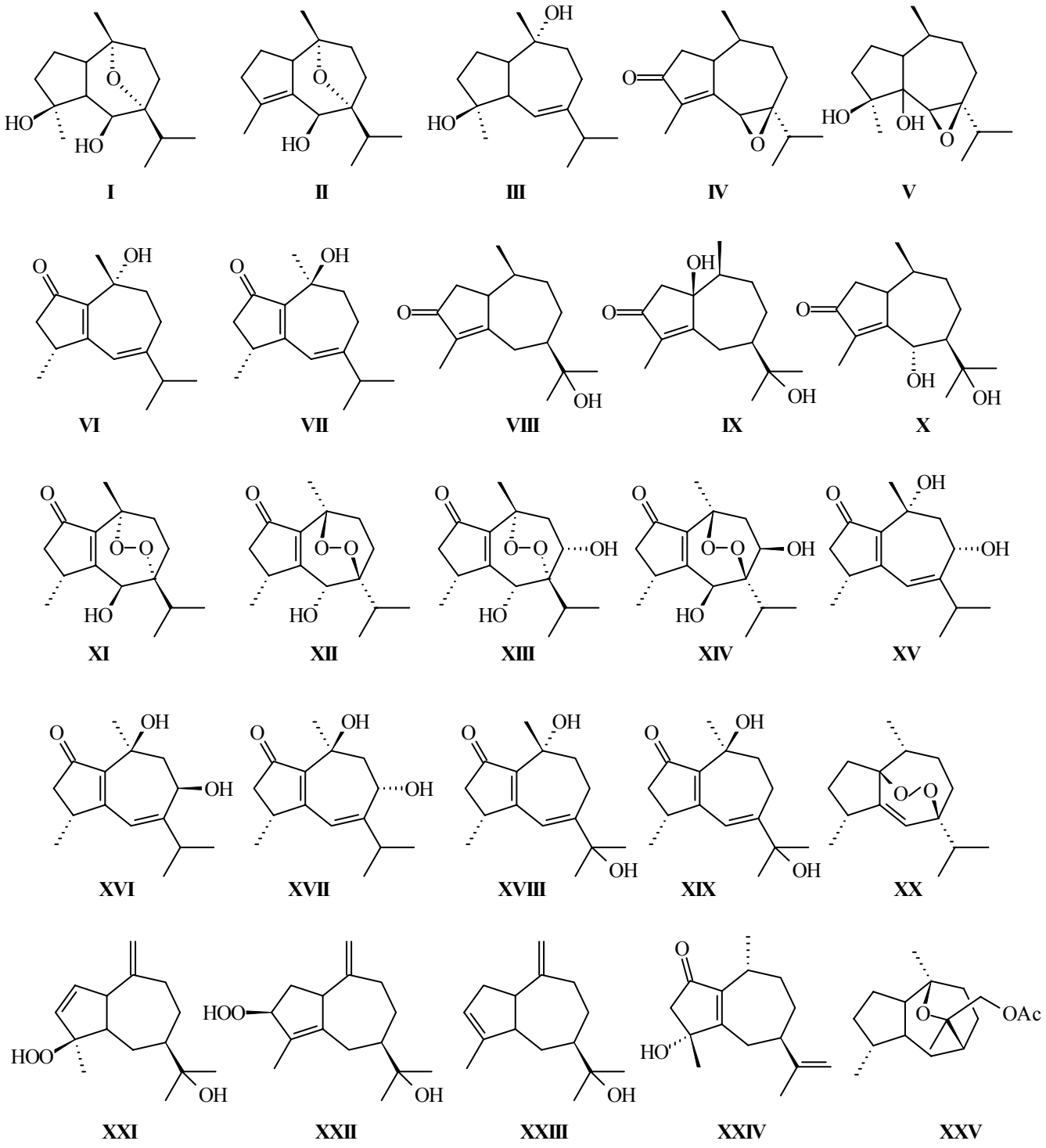

Figure 2. Substances used to test the system

Table 2. Substructures proposed by the system for guaianes in Figure 2.

Guaiane	^{13}C NMR data (C_1-C_{15})	Proposed substructures	References
I	50.4d, 23.5t, 40.1t, 79.1s, 54.5d, 72.1d, 86.4s, 29.0t, 31.8t, 83.0s, 32.4d, 17.4q, 18.3q, 25.1q, 23.6q	[4βOH; 14α] – 93.7% [6βOH; 7α,10αOXY] – 100.0%	[18]
II	57.7d, 24.0t, 39.1t, 133.6s, 133.0s, 74.0d, 86.7s, 28.6t, 31.7t, 84.5s, 31.8d, 17.3q, 18.2q, 14.7q, 24.1q	[6βOH; 7α,10αOXY] – 100.0%	[18]
III	50.9d, 21.6t, 40.6t, 80.3s, 50.0d, 121.4d, 149.7s, 25.2t, 42.5t, 76.8s, 37.3d, 21.2q, 21.5q, 22.7q, 21.5q	[4βOH; 14α] – 93.7% [6EN; 10αOH] – 100.0%	[18]
IV	43.3d, 40.5t, 208.5s, 141.1s, 167.9s, 59.7d, 70.1s, 20.9t, 29.0t, 32.8d, 36.9d, 18.1q, 18.1q, 10.2q, 7.9q	[3OXO; 4EN] – 100.0% [6β,7βEP] – 100.0%	[19]
V	45.7d, 26.4t, 33.2t, 69.5s, 69.1s, 64.3d, 66.2s, 21.1t, 29.0t, 33.9d, 36.9d, 18.2q, 18.1q, 12.6q, 15.2q	[6β,7βEP] – 100.0% [4βOH; 14α] – 93.7%	[19]
VI	141.0s, 210.2s, 43.3t, 36.8d, 167.5s, 117.4d, 166.1s, 27.3t, 38.9t, 72.3s, 38.6d, 21.3q, 21.0q, 20.4q, 28.8q	[1(5)EN; 2OXO] – 100.0% [1(5),6EN; 2OXO] –100.0% [6EN; 10αOH] – 100.0%	[20]
VII	141.2s, 210.0s, 43.3t, 36.8d, 167.5s, 117.6d, 166.0s, 27.8t, 38.7t, 71.9s, 38.5d, 21.6q, 21.1q, 20.2q, 27.9q	[1(5)EN; 2OXO] – 100.0% [1(5),6EN; 2OXO] –100.0% [6EN; 10OH] – 100.0%	[20]
VIII	45.7d, 41.3t, 208.8s, 137.7s, 176.4s, 33.7t, 47.9d, 27.0t, 36.7t, 35.3d, 73.1s, 27.2q, 25.9q, 12.1q, 7.9q	[3OXO; 4EN] – 100.0% [11OH] – 100.0%	[21]
IX	79.4s, 50.4t, 205.9s, 136.5s, 174.5s, 24.3t, 45.7d, 27.3t, 27.7t, 44.4d, 73.8s, 27.5q, 26.2q, 17.7q, 7.6q	[3OXO; 4EN] – 100.0% [1βOH; 3OXO; 4EN] – 97.0% [11OH] – 100.0%	[21]
X	42.9d, 40.4t, 210.5s, 139.2s, 173.9s, 70.0d, 53.1d, 21.9t, 34.2t, 33.6d, 74.0s, 28.7q, 23.3q, 13.7q, 7.2q	[3OXO; 4EN] – 100.0% [11OH] – 100.0%	[21]
XI	141.9s, 205.2s, 44.7t, 32.6d, 181.4s, 73.7d, 82.9s, 21.3t, 31.1t, 77.4s, 35.2d, 16.9q, 16.7q, 19.2q, 22.7q	[1(5)EN; 2OXO] – 100.0% [1(5)EN;6OH; 7,10Peroxy]–100.0% [6OH; 7,10Peroxy] – 100.0%	[22]
XII	140.9s, 205.5s, 45.9t, 36.2d, 178.9s, 75.4d, 83.8s, 20.1t, 33.9t, 77.1s, 36.3d, 17.4q, 16.7q, 20.4q, 23.0q	[1(5)EN; 2OXO] – 100.0% [1(5)EN;6OH; 7,10Peroxy]–100.0% [6OH; 7,10Peroxy] – 100.0%	[22]

Table 2. Continued

Guaiane	^{13}C NMR data (C_1-C_{15})	Proposed substructures	References
XIII	141.8s, 206.0s, 45.8t, 36.3d, 177.0s, 70.7d, 86.9s, 66.6d, 39.8t, 78.9s, 30.0d, 19.4q, 18.2q, 19.0q, 22.3q	[1(5)EN; 2OXO] – 100.0% [1(5)EN;6OH; 7,10Peroxy]–100.0% [6,8OH; 7,10Peroxy] – 100.0%	[22]
XIV	143.5s, 204.3s, 45.4t, 32.4d, 178.3s, 67.6d, 87.9s, 66.3d, 40.0t, 78.7s, 30.2d, 18.9q, 18.2q, 18.6q, 22.5q	[1(5)EN; 2OXO] – 100.0% [1(5)EN;6OH; 7,10Peroxy]–100.0% [6,8OH; 7,10Peroxy] – 100.0%	[22]
XV	141.6s, 210.5s, 43.0t, 36.7d, 169.9s, 113.1d, 167.9s, 69.0d, 54.2t, 71.2s, 30.3d, 22.2q, 22.3q, 20.4q, 28.2q	[1(5)EN; 2OXO] – 100.0% [1(5),6EN; 2OXO] –100.0% [6EN; 8,10OH] – 100.0%	[23]
XVI	141.8s, 210.6s, 43.3t, 36.8d, 169.3s, 114.0d, 167.9s, 69.0d, 53.9t, 70.9s, 30.1d, 22.3q, 22.5q, 20.0q, 28.0q	[1(5)EN; 2OXO] – 100.0% [1(5),6EN; 2OXO] –100.0% [6EN; 8,10OH] – 100.0%	[23]
XVII	140.3s, 209.9s, 43.3t, 36.8d, 167.1s, 117.4d, 166.4s, 68.8d, 45.8t, 71.5s, 35.7d, 21.4q, 21.1q, 19.9q, 28.2q	[1(5)EN; 2OXO] – 100.0% [1(5),6EN; 2OXO] –100.0% [6EN; 8,10OH] – 100.0%	[23]
XVIII	142.0s, 210.4s, 43.3t, 37.1d, 167.3s, 116.2d, 164.6s, 25.9t, 37.1t, ---s, 77.3s, 28.9q, 28.8q, 20.6q, 28.7q	[1(5)EN; 2OXO] – 100.0% [1(5),6EN; 2OXO] –100.0% [6EN; 10αOH] – 100.0%	[23]
XIX	142.1s, 210.4s, 43.4t, 37.0d, 167.1s, 116.5d, 164.4s, 26.1t, 39.7t, 72.0s, 74.2s, 29.1q, 28.7q, 20.2q, 28.2q	[1(5)EN; 2OXO] – 100.0% [1(5),6EN; 2OXO] –100.0% [6EN; 10OH] – 100.0%	[23]
XX	91.4s, 36.4t, 31.7t, 36.8d, 150.3s, 122.8d, 83.3s, 31.0t, 30.3t, 39.1d, 35.4d, 16.8q, 17.0q, 19.1q, 18.0q	---	[24]
XXI	54.1d, 133.6d, 136.6d, 96.5s, 48.8d, 27.4t, 48.7d, 27.3t, 39.4t, 153.4s, 73.9s, 26.4q, 27.8q, 106.9t, 19.2q	[10(14)EN; 11OH] – 100.0%	[25]
XXII	51.3d, 35.8t, 92.6d, 130.1s, 143.4s, 29.0t, 45.8d, 30.4t, 34.9t, 153.1s, 72.5s, 26.4q, 27.3q, 109.6t, 19.1q	[10(14)EN; 11OH] – 100.0%	[25]
XXIII	49.2d, 34.0t, 123.1d, 142.0s, 51.0d, 32.3t, 48.9d, 26.2t, 41.1t, 152.7s, 74.1s, 25.7q, 28.0q, 106.4t, 14.8q	[10(14)EN; 11OH] – 100.0%	[25]
XXIV	145.5s, 204.4s, 51.4t, 75.8s, 174.4s, 28.3t, 44.5d, 28.9t, 28.9t, 28.3d, 149.6s, 109.6t, 20.1q, 17.7q, 26.5q	[1(5)EN; 2OXO] – 100.0%	[26]
XXV	50.3d, 28.2t, 32.0t, 32.8d, 41.6d, 33.0t, 31.4d, 23.7t, 35.0t, 74.5s, 75.5s, 70.6t, 23.4q, 18.4q, 27.4q	[10,11OXY] – 100.0%	[27]

4 CONCLUSIONS

The overall goal of this study has been to develop an ability to predict the molecular structure of guaiane sesquiterpenes. In this case, the interplay between experimental measurements and computer-assisted determinations have led to a good understanding of the processes of carbon atom type determination and chemical shifts evaluation as previously described in the studied chemical structures. Accurate experimental determinations of chemical shifts have been used to test our computer-assisted elucidation process. Computational methods, validated by experiments, have then been used to examine the structural determination of terpenoids. The ^{13}C NMR rules obtained and exhibited in this work can be used by other research groups in the process of structure elucidation of new compounds.

In summary, we have shown here that substructures for guaiane sesquiterpenes can be predicted with 98.1% of accuracy by using the expert system SISTEMAT. The prediction requires as input data only the knowledge of experimental chemical shifts and the multiplicity of each carbon atom of the isolated substance structure. The use of the SISTEMAT might thus prove to be a convenient approach for predicting structures in the natural products' field. This paper shows that the predictive ability and applicability of a procedure based on ^{13}C NMR heuristic rules is strongly affected by the size of the learning database. In this sense our literature compilation has come closest to being complete in relation to the accurate experimental data available, although these are not very abundant. The results reported here for this new class of compounds, not previously presented in the SISTEMAT system, indicate and confirm the power of this expert system in the field of structural elucidation and suggest an efficient guide and a new tool for use in the substructure identification of sesquiterpenoids and in the search for ^{13}C NMR heuristic rules as well.

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. D. Connolly and R. A. Hill, *Dictionary of terpenoids*, Chapmann and Hall, London, 1991.
[2] I. W. Southon and J. Buckingham, *Dictionary of Alkaloids*, Eds. G. A. Cordell, J. E. Saxton, M. Shamma, G. F. Smith, Chapman and Hall, London, 1989.
[3] J. R. Hoult and M. Paya, Pharmacological and biochemical actions of simple coumarins: natural products with therapeutic potential, *Gen. Pharmacol.* **1996**, *27*, 713-722.
[4] M. J. Chan-Bacab and L. M. Peña-Rodríguez, Plant natural products with leishmanicidal activity, *Nat. Prod. Rep.* **2001**, *18*, 674-688.
[5] S. Schwikkard and F. R. van Heerden, Antimalarial activity of plant metabolites, *Nat. Prod. Rep.* **2002**, *19*, 675-692.
[6] P. M. Dewick, The biosynthesis of C_5-C_{25} terpenoid compounds, *Nat. Prod. Rep.* **2002**, *19*, 181-222.

[7] 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, *Química Nova* **1990**, *13*, 10-16.
[8] V. P. Emerenciano, G. V. Rodrigues, P. A. T. Macari, S. A. Vestri, J. H. G. Borges, J. P. Gastmans and D. L. G. Fromanteau, Applications d'intelligence artificielle dans la chimie organique. XVII. Nouveaux programmes du projet SISTEMAT, *Spectroscopy – An International Journal* **1994**, *12*, 91-98.
[9] M. E. Elyashberg, Y. Z. Karasev, E. R. Martirosian, H. Thiele and H. Somberg, Expert systems as a tool for the molecular structure elucidation by spectral methods. Strategies of solution to the problems, *Anal. Chim. Acta* **1997**, 348, 443-463.
[10] M. J. P. Ferreira, G. V. Rodrigues, A. J. C. Brant and V. P. Emerenciano, REGRAS: An auxiliary program for pattern recognition and substructure elucidation of monoterpenes, *Spectroscopy - An International Journal* **2001**, *15*, 65-98.
[11] M. J. P. Ferreira, G. V. Rodrigues and V. P. Emerenciano, MONOREG – an expert system for structural elucidation of monoterpenes, *Can. J. Chem.* **2001**, *79*, 1915-1925.
[12] F. C. Oliveira, M. J. P. Ferreira, C. V. Núñez, G. V. Rodrigues and V. P. Emerenciano, ^{13}C NMR spectroscopy of eudesmane sesquiterpenes, *Progress in NMR Spectroscopy* **2000**, *37*, 1-45.
[13] S. A. V. Alvarenga, J. P. Gastmans, G. V. Rodrigues and V. P. Emerenciano, Ditregra - an auxiliary program for structural determination of diterpenes, *Spectroscopy – An International Journal* **1997**, *13*, 227-249.
[14] P. A. T. Macari, J. P. Gastmans, G. V. Rodrigues and V. P. Emerenciano, An expert system for structure elucidation of triterpenes, *Spectroscopy – An International Journal* **1994**, *12*, 139-167.
[15] V. P. Emerenciano, L. D. Melo, G. V. Rodrigues and J. P. Gastmans, Application of Artificial Intelligence in Organic Chemistry. Part XIX. Pattern recognition and structural determination of flavonoids using ^{13}C-NMR spectra, *Spectroscopy – An International Journal* **1997**, *13*, 181-190.
[16] 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, *Progress in NMR Spectroscopy* **1998**, *33*, 153-206.
[17] M. J. P. Ferreira, F. C. Oliveira, S. A. V. Alvarenga, P. A. T. Macari, G. V. Rodrigues and V. P. Emerenciano, Automatic identification by ^{13}C NMR of substituents groups bonded in natural products skeletons, *Comp. & Chem.* **2002**, *26*, 601-632.
[18] G.-P. Peng, G. Tian, X.-F. Huang and F.-C. Lou, Guaiane-type sesquiterpenoids from *Alisma orientalis*, *Phytochemistry* **2003**, *63*, 877-881.
[19] R. Chowdhury, C. M. Hasan and M. A. Rashid, Guaiane sesquiterpenes from *Amoora rohituka*, *Phytochemistry* **2003**, *62*, 1213-1216.
[20] M. Tanitsu, Y. Takaya, M. Akasaka, M. Niwa and Y. Oshima, Guaiane- and aristolane-type sesquiterpenoids of *Nardostachys chinensis* roots, *Phytochemistry* **2002**, *59*, 845-849.
[21] C. Labbé, F. Faini, J. Coll and P. Carbonell, Guaiane sesquiterpenoids from *Haplopappus foliosus*, *Phytochemistry* **1998**, *49*, 793-795.
[22] Y. Takaya, Y. Takeuji, M. Akasaka, O. Nakagawasai, T. Tadano, K. Kisara, H.-S. Kim, Y. Wataya, M. Niwa and Y. Oshima, Novel guaiane endoperoxides, nardoguaianone A-D, from *Nardostachys chinensis* roots and their antinociceptive and antimalarial activities, *Tetrahedron* **2000**, *56*, 7673-7678.
[23] Y. Takaya, M. Akasaka, Y. Takeuji, M. Tanitsu, M. Niwa and Y. Oshima, Novel guaianoids, nardoguaianone E-I, from *Nardostachys chinensis* roots, *Tetrahedron* **2000**, *56*, 7679-7683.
[24] H. Hirota, T. Okino, E. Yoshimura and N. Fusetani, Five new antifouling sesquiterpenes from two marine sponges of the genus *Axinyssa* and the nudibranch *Phyllidia pustulosa*, *Tetrahedron* **1998**, *54*, 13971-13980.
[25] Y. Fukuyama, H. Minami, R. Ichikawa, K. Takeuchi and M. Kodama, Hydroperoxylated guaiane-type sesquiterpenes from *Viburnum awabuki*, *Phytochemistry* **1996**, *42*, 741-746.
[26] H. Achenbach and G. Benirschke, Joannesialactone and other compounds from *Joannesia princeps*, *Phytochemistry* **1997**, *45*, 149-157.
[27] M. Tori, M. Yoshida, M. Yokoyama, Y. Asakawa, A guaiane-type sesquiterpene, valeracetate from *Valeriana officinalis*, *Phytochemistry* **1996**, *41*, 977-979.