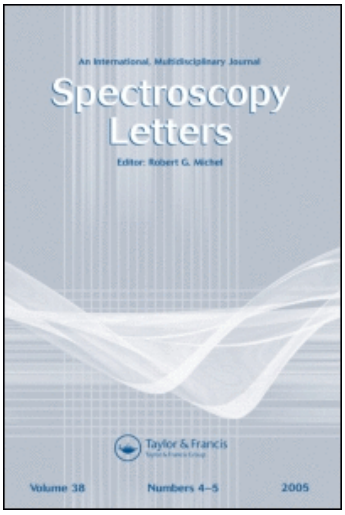

Spectroscopy Letters
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713597299

Application of a New Program, H1MACH, for Prediction of Iridoid Skeletons

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

Online publication date: 25 October 2004

To cite this Article Ferreira, Marcelo J. P. , Costantin, Mara B. , Rodrigues, Gilberto V. and Emerenciano, Vicente P.(2004) 'Application of a New Program, H1MACH, for Prediction of Iridoid Skeletons', Spectroscopy Letters, 37: 6, 587 — 605
To link to this Article: DOI: 10.1081/SL-200037615
URL: http://dx.doi.org/10.1081/SL-200037615

SPECTROSCOPY LETTERS
Vol. 37, No. 6, pp. 587–605, 2004

Application of a New Program, H1MACH, for Prediction of Iridoid Skeletons

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

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

ABSTRACT

A procedure is presented that utilizes ^{1}H NMR for prediction of the skeleton of iridoids. A new program was developed, named H1MACH, that presents a database with 800 data points from the ^{1}H NMR spectra of iridoids. This program was widely tested for the prediction of the skeleton of 40 compounds and compared with other programs in the expert system SISTEMAT. The results obtained show that H1MACH is very useful for the prediction of the skeleton of iridoids, especially for the iridane skeleton.

Key Words: Iridoids; Natural products; ^{1}H NMR; Skeleton prediction; Expert system.

*Correspondence: Vicente P. Emerenciano, Instituto de Química, Universidade de São Paulo, Caixa Postal 26077, 05513-970, São Paulo, Brazil; Fax: +55-11-38155579; E-mail: vdpemere@iq.usp.br.

DOI: 10.1081/SL-200037615 0038-7010 (Print); 1532-2289 (Online)

INTRODUCTION

The use of multispectral data, in expert systems whose objective is the structural determination of organic substances, is a well-known procedure. The DENDRAL system[1] was the pioneer in this area, and through data which originated from diverse spectroscopic sources, especially mass spectrometry, the system elucidated the structures of various organic substances. However, when the DENDRAL researchers tried to use the structure generator to identify chemical structures of natural products, whose structural diversity and complexity are higher, numerous problems were detected. The principal one among them was the combinatorial explosion in the structure generator. The solution to this problem was the use of information restrictions in the structural determination process, so that this process was not random. From this system, various others have been developed such as the ACCESS, DARC/EPIOS, SpecInfo, and, more recently, the Assemble 2.0 systems.[2–9]

In the last two decades, our research group has developed the expert system SISTEMAT,[10,11] main objective of which was the structural determination of natural products, particularly their carbonic skeletons. Various programs were developed that analyze spectral data with the objective to elucidate the carbonic skeleton of such compounds. The programs execute the analyses, for example, through disfunctionalization of ^{13}C NMR data;[12] identification of the skeletons and substructures by characteristic chemical shift ranges;[13–15] identification of the substituent groups bonded to natural product skeletons, through ^{13}C NMR data;[16] to analyze the botanical data, such as family and genera;[17,18] identification of the chemical constituents present in mixtures.[19]

The SISTEMAT system has been developed with the construction of specific databases of each natural product class, such as monoterpenes,[20] sesquiterpenes,[14] diterpenes,[15] triterpenes[21] and flavonoids.[22] Other classes such as steroids, lignoids, alkaloids, etc. are being inserted into the system.

The aim of this paper is to show and test a new program developed for the SISTEMAT expert system, the HIMACH program which predicts the skeleton of the compound from the ^{1}H NMR data analyses.

METHODOLOGY

For the development and testing of the H1MACH program, a database was created that contains the ^{1}H NMR data of a natural product class. The SISTEMAT, so far, had not been tested with ^{1}H NMR data from any

studied chemical class. Thus, the ^{1}H chemical shifts of the compounds were collected from the literature. The natural product class chosen for this study was the iridoids, and from this review was obtained 800 spectral data points from the ^{1}H NMR of these compounds.

The H1MACH Program

In the first attempt for building the database, we tried to link each chemical shift to the respective multiplicity as done in the ^{13}C NMR database. However, the signal multiplicity is strongly affected by the spectrometer type, 200, 300, 400, or 500 MHz, 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.

The H1MACH 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 percentual with the test sample. This percentual is computed like a Bremser's system.[2] After this attribution, the program selects the x-substances 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 can be varied from 1 to 50. To exemplify the use of the program, the iridoid shown in Fig. 1 was selected. After the input data, the program does the data match and shows the five compounds with the highest similarity index, Fig. 2. At the final step of the analysis, the program exhibits the skeleton probability for the compound tested. This probability (P) is calculated by: $P = (NS/TNS) \times 100$, 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

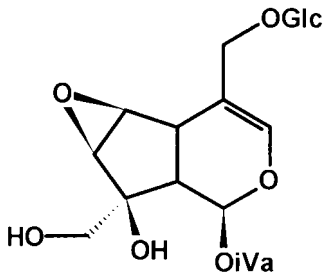

Figure 1. Iridoid employed to exemplify the H1MACH program.

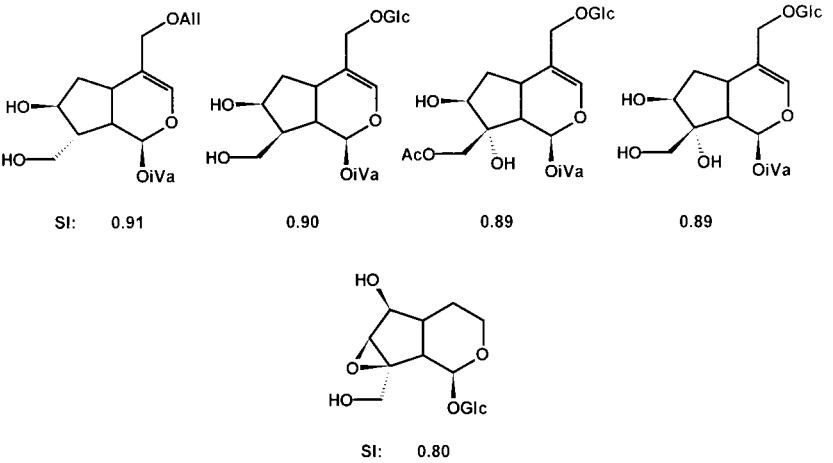

Figure 2. Iridoids with higher similarity index selected by the H1MACH program.

question is computed. For the iridoid in Fig. 1, the skeleton probability is shown in Table 1.

RESULTS

In order to test the action of the H1MACH program and to evaluate its efficiency, the ^{1}H NMR spectra data of 40 iridoids were randomly selected from the literature.[23–39] Also, the same compounds were tested with the C13MACH program[20,21] that did the same type of analysis but for ^{13}C NMR spectra data. The structures of the compounds used to test both programs are shown in Fig. 3.

The results obtained by use of the programs H1MACH and C13MACH are shown in Table 2. This contains the ^{1}H NMR and ^{13}C NMR data of the respective substance, the three most probable skeletons proposed by the programs H1MACH and C13MACH and the respective references. Figure 4 shows the structures of the skeletons proposed by the program.

Table 1. Results presented by the H1MACH program.

Skeletal type	Probability
Iridane	90.48
11-Nor-iridane	9.52

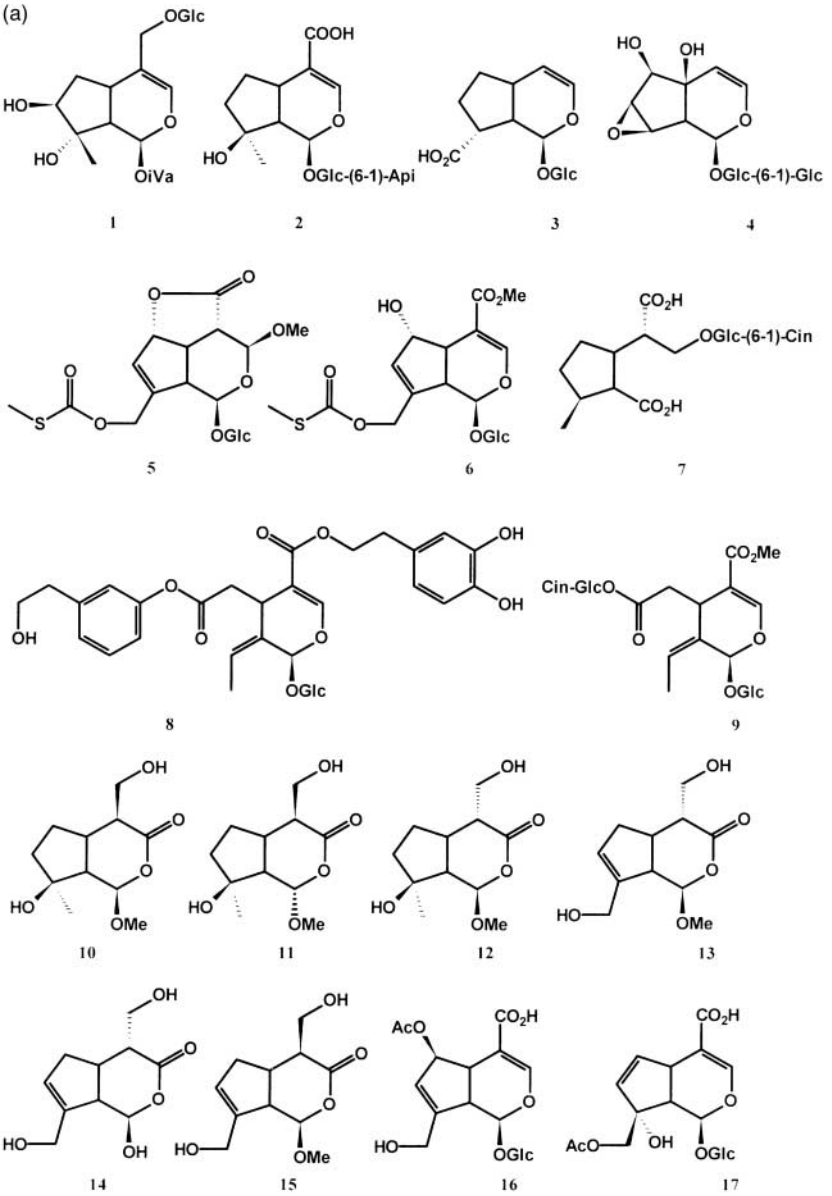

Figure 3. Substances used to test the programs.

(continued)

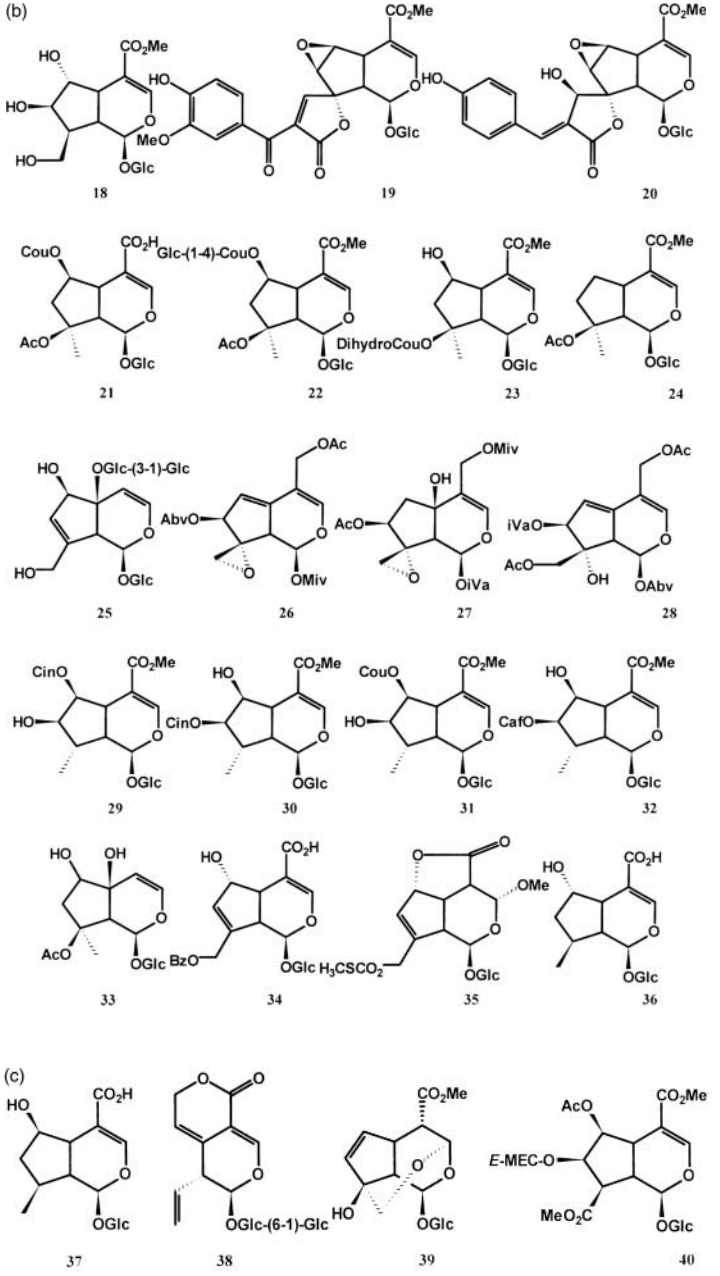

Figure 3. Continued.

Table 2. Chemical shifts of the substances and Results obtained by the C13MACH and H1MACH programs.

Substance	^{1}H NMR data	H1MACH program skeletal type (%)	^{13}C NMR data	C13MACH program Skeletal type (%)[a]	References
1^{M}	6.39, 6.43, 3.07, 4.05, 3.35, 2.01, 3.68, 3.68, 4.22, 4.33	*Iridane*: 90.48, 11nor-iridane: 9.52	90.6, 142.7, 109.2, 35.5, 59.7, 60.2, 80.1, 43.4, 67.0, 69.6	*Iridane*: 69.3, 11nor-iridane: 30.7	[23]
2^{M}	5.34, 7.40, 3.16, 1.46, 1.71, 1.71, 2.18, 1.34, 1.34, 1.34	*Iridane*: 100.0	95.6, 152.1, 113.5, 32.5, 30.8, 40.5, 80.4, 52.3, 24.8, 170.8	*Iridane*: 67.3, Skeleton-VII: 32.7	[24]
3^{M}	5.26, 6.21, 4.72, 2.78, 1.96, 1.49, 2.05, 1.84, 2.80, 2.47	*11nor-iridane*: 59.45, 10,11dinor-iridane: 21.40, Iridane: 19.02	95.5, 140.7, 107.6, 34.6, 33.0, 29.0, 45.8, 46.4, 179.5	Iridane: 79.7, 10nor-iridane: 10.6, *11nor-iridane*: 9.7	[25]
4^{M}	5.21, 6.37, 4.92, 4.06, 3.53, 3.62, 2.45	11nor-iridane: 50.00, *10,11dinor-iridane*: 40.00, 10nor-iridane: 10.00	96.5, 143.1, 108.2, 73.6, 79.2, 59.4, 56.3, 50.7	10nor-iridane: 49.2, 11nor-iridane: 37.7, *10,11dinor-iridane*: 13.1	[25]
5^{M}	5.10, 5.01, 3.25, 3.40, 5.37, 5.98, 3.02, 4.89, 5.08	11nor-iridane: 81.58, 10,11dinor-iridane: 18.42, *iridane*: 0.00	96.7, 98.5, 44.4, 37.6, 87.7, 126.8, 151.6, 46.3, 65.2, 177.0	*Iridane*: 41.2, 11nor-iridane: 29.0, 7,8seco-iridane: 20.2	[26]
6^{M}	5.06, 7.65, 3.03, 4.80, 6.02, 2.62, 4.95, 5.10	*Iridane*: 98.73, 7,8seco-iridane: 1.27	101.3, 155.4, 108.1, 42.4, 75.3, 132.4, 145.5, 46.2, 66.2, 172.9	*Iridane*: 67.2, 11nor-iridane: 16.6, 7,8seco-iridane: 8.2	[26]

(*continued*)

Table 2. Continued.

Substance	^{1}H NMR data	H1MACH program skeletal type (%)	^{13}C NMR data	C13MACH program Skeletal type (%)[a]	References
7^{M}	3.70, 4.12, 2.84, 2.51, 1.53, 1.85, 1.14, 1.93, 2.25, 2.51, 1.02, 1.02, 1.02	11nor-iridane: 82.30, *Iridane*: 17.60	179.2, 71.3, 49.2, 41.9, 30.8, 34.5, 40.1, 55.3, 22.0, 178.1	*Iridane*: 61.5, 7,8seco-iridane: 29.7, 10,11dinor-iridane: 9.4	[27]
8^{M}	6.01, 7.52, 4.07, 2.66, 2.87, 6.16, 1.74, 1.74, 1.74	Iridane: 97.35, *7,8seco-iridane*: 2.65	95.3, 155.3, 109.5, 31.8, 41.0, 171.7, 125.1, 130.6, 13.8, 168.2	*7,8seco-iridane*: 89.7, skeleton VII: 10.3	[27]
9^{M}	5.91, 7.51, 4.00, 2.50, 2.75, 6.09, 1.73, 1.73, 1.73	Iridane: 60.00, *7,8seco-iridane*: 40.00	95.2, 155.2, 109.4, 31.8, 41.2, 173.0, 125.0, 130.5, 13.7, 168.8	*7,8seco-iridane*: 60.0, skeleton VII: 40.0	[27]
10	5.07, 2.00, 3.10, 1.75, 1.97, 1.70, 1.97, 2.69, 1.23, 1.23, 1.23, 3.24, 3.85	*Iridane*: 66.67, 11nor-iridane: 33.33	92.7, 173.0, 49.3, 38.3, 23.2, 37.9, 80.1, 48.2, 24.1, 63.7	*Iridane*: 50.6, 11nor-iridane: 40.0, skeleton IV: 9.4	[28]
11	5.16, 2.10, 3.09, 1.60, 2.14, 1.79, 1.83, 2.63, 1.28, 1.28, 1.28, 3.48, 4.06	*Iridane*: 71.43, 11nor-iridane: 28.57	91.7, 172.9, 46.5, 34.8, 27.1, 38.9, 80.1, 46.8, 25.0, 61.4, 51.9	*Iridane*: 52.7, skeleton II: 19.1, 11nor-iridane: 18.7	[28]

12	5.00, 2.06, 2.61, 1.61, 1.98, 1.75, 1.76, 1.80, 2.39, 2.39, 2.39, 1.35, 3.73	*Iridane*: 53.85, 11nor-iridane: 46.15	94.6, 173.6, 47.7, 38.3, 27.3, 40.2, 79.7, 49.9, 25.5, 61.0, 51.0	*Iridane*: 49.4, 11nor-iridane: 40.8, 10,11dinor-iridane: 9.8	[28]
13	4.81, 2.33, 2.59, 2.01, 2.70, 5.70, 2.36, 4.20, 4.20, 3.49, 3.95	11nor-iridane: 52.48, *iridane*: 47.52	95.2, 173.9, 43.6, 39.6, 35.9, 128.4, 143.1, 50.9, 60.1, 63.6	*Iridane*: 51.6, 11nor-iridane: 38.7, skeleton VII: 9.7	[28]
14	4.19, 2.37, 2.80, 2.18, 2.22, 5.73, 2.38, 4.25, 4.25, 3.46, 4.00	*Iridane*: 60.81, 11nor-iridane: 39.19	92.9, 173.7, 39.5, 38.3, 29.7, 127.2, 143.2, 50.9, 60.3, 63.7	11nor-iridane: 61.0, *iridane*: 29.3, 7,8seco-iridane: 9.7	[28]
15	4.46, 2.99, 2.91, 2.19, 2.27, 5.81, 2.41, 4.23, 4.23, 3.71, 4.10	11nor-iridane: 81.44, *iridane*: 18.56	98.7, 172.4, 40.9, 40.5, 30.9, 128.7, 144.8, 50.7, 61.4, 62.3	*Iridane*: 60.11, 1nor-iridane: 39.9	[28]
16^{M}	5.23, 7.46, 3.26, 5.56, 5.75, 3.00, 4.32, 4.15	*Iridane*: 78.20, skeleton II: 11.60, 11nor-iridane: 10.10	97.7, 154.0, 110.1, 42.1, 83.6, 127.0, 150.3, 46.8, 60.9, 170.1	*Iridane*: 69.1, 11nor-iridane: 20.0, Skeleton I: 10.8	[29]
17^{M}	5.64, 7.36, 3.28, 6.22, 5.60, 2.61, 4.16, 4.06	*Iridane*: 79.50, skeleton II: 20.50	94.7, 152.4, 111.0, 39.0, 138.2, 132.7, 84.1, 46.5, 70.6, 170.1	*Iridane*: 62.9, 7,8seco-iridane: 37.1	[29]

(*continued*)

Table 2. Continued.

Substance	^{1}H NMR data	H1MACH program skeletal type (%)	^{13}C NMR data	C13MACH program Skeletal type (%)[a]	References
18^{M}	5.03, 7.58, 3.08, 4.22, 4.05, 2.32, 1.80, 3.79, 3.75	*Iridane*: 76.50, 11nor-iridane: 18.50, 7,8seco-iridane: 5.00	102.4, 155.7, 107.4, 40.6, 79.3, 77.5, 48.0, 40.5, 62.0, 169.4	11nor-iridane: 49.6, *iridane*: 31.2, skeleton IV: 9.7	[29]
19^{M}	5.60, 7.51, 3.46, 4.02, 3.50, 2.85, 7.56, 7.42, 6.84, 7.45	Iridane: 86.21, 11nor-iridane: 13.26, 7,8seco-iridane: 0.53, *skeleton III*: 0.00	92.7, 153.6, 108.5, 33.1, 58.0, 59.2, 92.7, 44.2, 156.2, 133.9, 169.3, 187.7, 167.8, 129.1, 113.0, 149.2, 154.8, 116.2, 126.8	11nor-iridane: 70.2, Iridane: 29.8	[29]
20^{M}	5.35, 7.43, 3.37, 4.04, 3.83, 2.45, 5.13, 7.57, 3.71, 7.61, 6.84, 6.84, 7.61	11nor-iridane: 79.00, skeleton III: 21.00, *skeleton III*: 0.00	92.8, 153.3, 108.0, 33.2, 58.1, 58.2, 92.6, 45.1, 69.1, 123.9, 172.7, 144.0, 168.0, 126.2, 134.8, 117.0, 162.1, 117.0, 134.8	7,8seco-iridane: 50.1, 11nor-iridane: 29.8, iridane: 20.1	[29]
21^{M}	5.89, 7.49, 3.33, 5.40, 2.37, 2.05, 3.01, 1.51, 1.51, 1.51, 1.51	*Iridane*: 100.00	95.3, 154.6, 108.4, 39.8, 78.8, 45.2, 89.6, 50.3, 21.9, 169.7	*Iridane*: 90.5, 7,8seco-iridane: 9.5	[30]

22[M]	5.85, 7.48, 3.30, 5.34, 2.38, 2.09, 3.00, 1.53, 1.53, 1.53	*Iridane*: 100.00	95.4, 154.5, 108.5, 39.9, 78.9, 45.1, 89.6, 50.3, 21.8, 168.4	*Iridane*: 58.5, 7,8seco-iridane: 32.0, skeleton IV: 9.5	[30]
23[M]	5.82, 7.38, 2.94, 4.25, 2.11, 1.97, 2.94, 1.38, 1.38, 1.38	*Iridane*: 100.00	95.7, 153.6, 109.8, 42.2, 76.0, 47.7, 89.7, 49.9, 22.2, 169.0	*Iridane*: 50.0, 7,8seco-iridane: 39.8, 11nor-iridane: 10.1	[30]
24[M]	5.71, 7.43, 3.12, 1.75, 1.75, 2.05, 2.05, 2.68, 1.54, 1.54, 1.54	*Iridane*: 100.00	95.5, 153.0, 112.2, 32.9, 29.7, 39.6, 91.0, 51.0, 21.2, 169.0	*Iridane*: 81.8, 7,8seco-iridane: 9.2, skeleton V: 9.0	[30]
25[M]	5.63, 6.65, 5.15, 5.79	*11nor-iridane*: 100.00	93.8, 143.4, 105.0, 79.8, 79.6, 128.0, 147.6, 51.4, 60.9	*11nor-iridane*: 100.0	[31]
26	5.98, 6.70, 5.85, 5.38, 3.44, 2.91, 3.02, 4.66, 4.76	*Iridane*: 60.50, 11nor-iridane: 39.50	92.6, 148.6, 108.5, 141.1, 118.6, 83.6, 64.2, 43.3, 48.0, 61.0	*Iridane*: 100.0	[32]
27	6.08, 6.68, 1 .99, 2.79, 4.87, 2.91, 2.82, 3.12, 4.72, 4.90	*Iridane*: 89.90, 11nor-iridane: 10.10	88.5, 145.4, 111.3, 69.7, 40.4, 73.3, 62.4, 48.2, 48.9, 61.9	*Iridane*: 91.2, skeleton VI: 8.8	[32]
28	6.20, 6.68, 5.74, 5.37, 2.89, 4.25, 4.68, 4.60, 4.71, 2.80, 3.04	*Iridane*: 100.0	92.7, 148.4, 108.5, 139.2, 117.3, 83.2, 80.3, 48.5, 65.8, 60.9	*Iridane*: 100.0	[32]

(*continued*)

Table 2. Continued.

Substance	^{1}H NMR data	H1MACH program skeletal type (%)	^{13}C NMR data	C13MACH program Skeletal type (%)[a]	References
29[A]	5.58, 7.43, 3.04, 5.40, 3.81, 2.35, 2.82, 1.13, 1.13, 1.13	*Iridane*: 99.89, 7,8seco-iridane: 0.11	95.0, 152.9, 109.7, 36.8, 78.8, 78.1, 39.8, 40.0, 14.0, 166.9	*Iridane*: 70.5, 11nor-iridane: 29.5	[33]
30[A]	5.61, 7.43, 2.96, 4.34, 4.72, 2.67, 2.88, 1.11, 1.11, 1.11	*Iridane*: 85.80, 11nor-iridane: 14.20	95.0, 152.9, 110.2, 39.0, 74.6, 81.6, 36.2, 39.2, 13.9, 167.5	*Iridane*: 40.9, 11nor-iridane: 28.6, 7,8seco-iridane: 19.9	[33]
31[A]	5.58, 7.42, 2.99, 5.38, 3.79, 2.34, 2.81, 1.12, 1.12, 1.12	*Iridane*: 99.90, 7,8seco-iridane: 0.10	95.0, 152.9, 109.8, 36.8, 78.5, 78.2, 39.8, 40.0, 14.0, 166.9	*Iridane*: 90.1, 7,8seco-iridane: 9.9	[33]
32[A]	5.59, 7.44, 2.94, 4.32, 4.69, 2.64, 2.85, 1.09, 1.09, 1.09	*Iridane*: 85.80, 11nor-iridane: 13.60, 7,8seco-iridane: 0.60	95.0, 152.8, 110.3, 39.0, 74.7, 81.3, 36.2, 39.3, 13.9, 167.4	*Iridane*: 40.1, 7,8seco-iridane: 30.1, 11nor-iridane: 19.3	[33]
33[W]	6.08, 6.46, 5.01, 3.84, 2.17, 2.03, 2.86, 1.45, 1.45, 1.45	*11nor-iridane*: 51.80, Iridane: 48.20	96.6, 145.3, 107.6, 75.0, 79.0, 47.0, 90.6, 55.8, 24.0	*11nor-iridane*: 53.7, iridane: 27.1, 10,11dinor-iridane: 10.2	[34]
34[M]	5.11, 7.63, 3 .08, 6.10, 2.71, 5.24, 5.04	11nor-iridane: 61.35, *Iridane*: 35.79, 7,8seco-iridane: 2.86	101.3, 154.8, 109.0, 42.7, 75.5, 132.0, 146.0, 46.6, 64.3, 170.0	*Iridane*: 82.6, 11nor-iridane: 8.8, skeleton VI: 8.6	[35]

35^{M}	5.09, 5.01, 3.25, 3.40, 5.37, 5.98, 3.00, 5.08, 4.90	11nor-iridane: 81.58, *Iridane*: 18.42	96.7, 98.6, 44.4, 37.7, 87.7, 126.8, 151.6, 46.3, 65.2, 177.0	11nor-iridane: 51.0, *iridane*: 29.0, skeleton VII: 10.3	[35]
36^{M}	5.21, 7.62, 2.82, 4.47, 1.38, 1.92, 2.30, 1.70, 1.12, 1.12, 1.12	*Iridane*: 100.00	101.2, 155.9, 107.4, 43.5, 75.1, 43.2, 35.2, 47.0, 21.9, 171.1	*Iridane*: 70.8, 11nor-iridane: 29.2	[36]
37^{M}	5.25, 7.41, 2.79, 4.05, 1.25, 2.17, 1.96, 2.03, 1.15, 1.15, 1.15	*iridane*: 100.00	97.5, 153.6, 110.8, 43.7, 78.8, 42.7, 34.3, 47.9, 21.1, 171.0	*Iridane*: 80.2, 7,8seco-iridane: 19.8	[36]
38^{M}	5.64, 7.45, 5.61, 5.07, 5.00, 5.76, 3.30, 5.24, 5.22	*7,8seco-iridane*: 40.00, 11nor-iridane: 20.00, iridane: 10.00	98.8, 150.8, 105.0, 127.1, 117.2, 70.9, 135.0, 46.7, 118.8, 166.4	*7,8seco-iridane*: 50.2, iridane: 25.6, 11nor-iridane: 8.6	[37]
39^{M}	5.69, 5.54, 3.59, 3.35, 6.10, 5.63, 2.52, 3.78, 3.75	*Iridane*: 50.00, 11nor-iridane: 41.25, 10,11dinor-iridane: 8.75	94.0, 95.7, 49.6, 38.0, 135.2, 137.4, 84.5, 53.2, 67.1, 172.4	*Iridane*: 79.9, skeleton IV: 10.6, 11nor-iridane: 9.5	[38]
40^{M}	5.39, 7.53, 3.30, 5.23, 5.70, 3.30, 3.09	*iridane*: 100.00	97.2, 154.1, 109.5, 37.4, 78.7, 73.7, 48.1, 39.8, 171.8, 168.7	*Iridane*: 47.5, 7,8seco-iridane: 42.8, skeleton VI: 9.6	[39]

Note: In italics represents the correct skeleton of the substance. Solvents: $CDCl_3$; M = CD_3OD; A = Acetone-d^6; and W = D_2O.

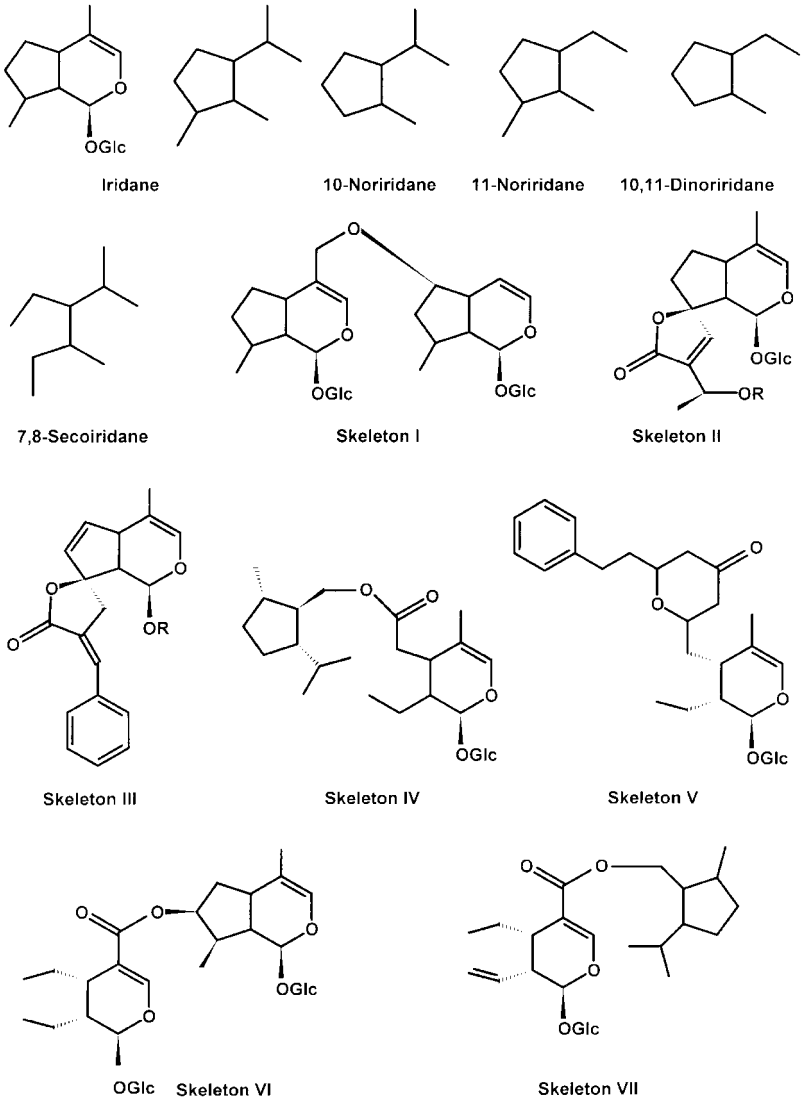

Figure 4. Skeletons proposed by the H1MACH program.

DISCUSSION OF THE RESULTS

The analysis of 40 iridoids was done through the programs H1MACH and C13MACH. The program H1MACH showed 72.50% accuracy. In the 11 incorrect proposals of skeleton, one can verify the following.

(1) In tests 4, 7, 8, 9, 13, 15, 34, and 35, the program supplies the correct skeleton as the second more probable skeleton.

(2) In tests 5, 19, and 20, the correct skeleton was not proposed by the program.

Analyzing the incorrect proposals in a more detailed way, one can verify for the tests 4, 5, 7, 13, 15, 34, and 35 that the program furnished the corresponding nor-derivatives for the correct skeleton. This error was due to the higher structure similarity between the tested substance and the database substances pertaining to a nor-skeleton. The only existing difference between the tested substances and the selected substances is the chemical shift referring to the methyl group, which was insufficient to differentiate between the skeletons. It is also necessary to comment that in tests 19 and 20 it is impossible for the program to predict the correct skeleton of the substances because the database does not contain the ^{1}H NMR spectral data of the respective skeletons.

The program C13MACH, that executes ^{13}C NMR analysis, displayed the correct skeleton of the substance in 82.50% of the cases. In the seven incorrect proposals for the carbon skeletons, it was verified that the correct skeletons were the second in three cases (tests 14, 18, and 35) and the third in two cases (tests 3 and 4). In the two last cases (tests 19 and 20), the correct skeleton was not exhibited; this error was due to presence of few ^{13}C NMR spectra, and only one of the correct skeletons in the database.

CONCLUSIONS

Regarding the results obtained, it can be concluded that the tests done with the program H1MACH showed good results, once the signal multiplicity was not included in the database. For the cases where the program mistakes the skeleton prediction it can be observed that the correct skeleton of the substance is found among the three first skeletons proposed by the program in 92.50% of the cases. In the future, the program H1MACH might be utilized as a restriction module for the structure generator that is already being developed for the expert system SISTEMAT. This program will be integrated into a program set of the latter system, which should display only one skeleton

proposal based on the sum of the weight probabilities of each program. Thus, this final probability will be used to select skeletons as the starting point for the structure generator. Therefore, instead of the generator working randomly to start the process of generation of likely structures, it will have to initiate the process by using only the three first skeletons proposed by the programs. The immediate consequence of the use of this novel program in the structure generator will be the reduction of the computational time spent, and the number of displayed candidate structures, which avoids the combinatorial explosion problem observed in other expert systems developed up to now.

The results obtained with the C13MACH program were a little higher than the H1MACH results, because for ^{13}C NMR data the signal multiplicity was inserted in the system. The C13MACH program executes simple data matching. However, when the SISCONST program[40] was used, the results were higher than the ones of the latter program. The SISCONST program matches each signal of the spectral data with the ones stored in the database, and if a signal and its multiplicity are present in a determined carbon atom, the signals of the interlinked carbons are matched with the data in question. This searching process is repeated so that the largest fragments of the substructure bearing compatible chemical shifts with the ^{13}C NMR data from the spectrum are obtained, thus the SISCONST results are better.

ACKNOWLEDGMENTS

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).

REFERENCES

1. Lindsay, R.K.; Buchanan, B.G.; Feigenbaum, E.A.; Lederberg, J. *Applications of Artificial Intelligence for Organic Chemistry: The DENDRAL Project*, 1st Ed.; McGraw-Hill: New York, 1980; Vol. 1, 1–189.
2. Bremser, W. Multidimensional spectroscopy. Magn. Reson. Chem. **1985**, *23* (12), 1056–1071.
3. 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.
4. 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.
5. 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.
6. 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.
7. 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.
8. Munk, M.E. Computer-based structure determination: then and now. J. Chem. Inf. Comput. Sci. **1998**, *38* (6), 997–1009.
9. 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.
10. 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.
11. Emerenciano, V.P.; Rodrigues, G.V.; Macari, P.A.T.; Vestri, S.A.; Borges, J.H.G.; Gastmans, J.P.; Fromanteau, D.L.G. Applications d'intelligence artificielle dans la chimie organique. XVII. Nouveaux programmes du projet SISTEMAT. Spectroscopy **1994**, *12* (2), 91–98.
12. Ferreira, M.J.P.; Brant, A.J.C.; Rodrigues, G.V.; Emerenciano, V.P. Automatic identification of terpenoid skeletons through ^{13}C NMR data disfuncionalization. Anal. Chim. Acta **2001**, *429* (1), 151–170.
13. Ferreira, M.J.P.; Rodrigues, G.V.; Brant, A.J.C.; Emerenciano, V.P. REGRAS: an auxiliary program for pattern recognition and substructure elucidation of monoterpenes. Spectroscopy **2001**, *15* (2), 65–98.
14. 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.
15. 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.
16. 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. **2002**, *26* (6), 601–632.

17. 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.
18. Alvarenga, S.A.V.; Gastmans, J.P.; Ferreira, M.J.P.; Rodrigues, G.V.; Brant, A.J.C.; Emerenciano, V.P. SISTAX—an intelligent tool for recovering information on natural products chemistry. J. Braz. Chem. Soc. **2003**, *14* (3), 369–374.
19. Ferreira, M.J.P.; Costantin, M.B.; Sartorelli, P.; Rodrigues, G.V.; Limberger, R.; Henriques, A.T.; Kato, M.J.; Emerenciano, V.P. A computer-aided method for individual components identification by ^{13}C NMR of essential oils. Anal. Chim. Acta **2001**, *447* (1–2), 125–134.
20. 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.
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.; Melo, L.D.; Rodrigues, G.V.; Gastmans, J.P. Application of artificial intelligence in organic chemistry. Part XIX. Pattern recognition and structural determination of flavonoids using ^{13}C-NMR spectra. Spectroscopy **1997**, *13* (3), 181–190.
23. Kuruüzüm-Uz, A.; Güvenalp, Z.; Demirezer, L.Ö.; Bergère, I.; Ströch, K.; Zeeck, A. 4′-Deoxy iridoid glycosides from *Centranthus longiflorus*. Phytochemistry **2002**, *61* (8), 937–941.
24. Kanchanapoom, T.; Kasai, R.; Yamasaki, K. Iridoid and phenolic diglycosides from *Canthium berberidifolium*. Phytochemistry **2002**, *61* (4), 461–464.
25. Kanchanapoom, T.; Kasai, R.; Yamasaki, K. Iridoid glycosides from *Thunbergia laurifolia*. Phytochemistry **2002**, *60* (8), 769–771.
26. Quang, D.N.; Hashimoto, T.; Tanaka, M.; Dung, N.X.; Asakawa, Y. Iridoid glucosides from roots of Vietnamese *Paederia scandens*. Phytochemistry **2002**, *60* (5), 505–514.
27. Takenaka, Y.; Okazaki, N.; Tanahashi, T.; Nagakura, N.; Nishi, T. Seco-iridoid and iridoid glucosides from *Syringa afghanica*. Phytochemistry **2002**, *59* (7), 779–787.
28. Daí, J.-Q.; Liu, Z.-L.; Yang, L. Non-glycosidic iridoids from *Cymbaria mongolica*. Phytochemistry **2002**, *59* (5), 537–542.
29. Kanchanapoom, T.; Kasai, R.; Yamasaki, K. Iridoid and phenolic glycosides from *Morinda coreia*. Phytochemistry **2002**, *59* (5), 551–556.
30. Kanchanapoom, T.; Kasai, R.; Yamasaki, K. Iridoid glucosides from *Barleria lupulina*. Phytochemistry **2001**, *58* (2), 337–341.

31. Kanchanapoom, T.; Kasai, R.; Chumsri, P.; Hiraga, Y.; Yamasaki, K. Megastigmane and iridoid glucosides from *Clerodendrum inerme*. Phytochemistry **2001**, *58* (2), 333–336.
32. Tang, Y.; Liu, X.; Yu, B. Iridoids from the rhizomes and roots of *Valeriana jatamansi*. J. Nat. Prod. **2002**, *65* (12), 1949–1952.
33. Ayers, S.; Sneden, A.T. Caudatosides A-F: new iridoid glucosides from *Citharexylum caudatum*. J. Nat. Prod. **2002**, *65* (11), 1621–1626.
34. Kuria, K.A.M.; Chepkwony, H.; Govaerts, C.; Roets, E.; Busson, R.; de Witte, P.; Zupko, I.; Hoornaert, G.; Quirynen, L.; Maes, L.; Janssens, L.; Hoogmartens, J.; Laekeman, G. The antiplasmodial activity of isolates from *Ajuga remota*. J. Nat. Prod. **2002**, *65* (5), 789–793.
35. Ling, S.; Komorita, A.; Tanaka, T.; Fujioka, T.; Mihashi, K.; Kouno, I. Sulfur-containing bis-iridoid glucosides and iridoid glucosides from *Saprosma scortechinii*. J. Nat. Prod. **2002**, *65* (5), 656–660.
36. Tanaka, N.; Tanaka, T.; Fujioka, T.; Fujii, H.; Mihashi, K.; Shimomura, K.; Ishimaru, K. An ellagic compound and iridoids from *Cornus capitata* root cultures. Phytochemistry **2002**, *57* (8), 1287–1291.
37. Kakuda, R.; Iijima, T.; Yaoita, Y.; Machida, K.; Kikuchi, M. Secoiridoid glycosides from *Gentiana scabra*. J. Nat. Prod. **2001**, *64* (12), 1574–1575.
38. Ling, S.; Tanaka, T.; Kouno, I. Iridoids from *Rothmannia macrophylla*. J. Nat. Prod. **2001**, *64* (6), 796–798.
39. Franzyk, H.; Jensen, S.R.; Olsen, C.E. Iridoid glucosides from *Myxopyrum smilacifolium*. J. Nat. Prod. **2001**, *64* (5), 632–633.
40. 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. **1993**, *17* (4), 369–378.

Received February 14, 2004
Accepted August 20, 2004