Ditregra – an auxiliary program for structural determination of diterpenes*

Sandra A. Vestri Alvarenga^a, Jean Pierre Gastmans^a, Gilberto do Vale Rodrigues^b and Vicente de Paulo Emerenciano^b

^a*Faculdade de Engenharia, Guaratingueta-UNESP, Av. Dr Ariberto Pereira da Cunha 333, Guaratingueta, Sao Paulo, Brazil*

^b*Instituto de Química-USP, Av. Lineu Prestes 748, São Paulo, Brazil, C. P. 26.077, CEP: 05599-970*

Abstract. This work describes the creation of heuristics rules based on ¹³C-NMR spectroscopy that characterize several skeletal types of diterpenes. Using a collection of 2745 spectra we built a database linked to the expert system SISTEMAT. Several programs were applied to the database in order to discover characteristic signals that identify with a good performance, a large diversity of skeletal types. The heuristic approach used was able to differentiate groups of skeletons based firstly on the number of primary, secondary, tertiary and quaternary carbons, and secondly the program searches, for each group, if there are ranges of chemical shifts that identifies specific skeletal type. The program was checked with 100 new structures recently published and was able to identify the correct skeleton in 65 of the studied cases. When the skeleton has several hundreds of compounds, for example, the labdanes, the program employs the concept of subskeletal, and does not classify in the same group labdanes with double bounds at different positions. The chemical shift ranges for each subskeletal types and the structures of all skeletal types are given. The consultation program can be obtained from the authors.

1. Introduction

There exist specialised systems developed to assist the chemist in structure elucidation work. These systems try to imitate the thinking process followed by the chemist when using various spectral data in order to arrive at a proposed substructure. These substructures are accompanied by a program that generates and provides a list of spectral propositions. Following this, some proposals are rejected throughout the analysis, and other types of information are gathered by comparing theoretical spectra with experimental ones, results obtained from synthesis, etc.

The major systems that operate this way are DENDRAL [9], DARC-EPIOS [1] and CASE [10]. In order to reduce the number of proposed structures, these systems generate restrictions which are substructures (fragments of structures). For large molecules (with more than 15 atoms), the structural fragments must contain a large number of atoms in order to avoid a combinatory explosion and a list of proposed structures that would be too large. Recently some systems have introduced 2D-Nuclear Magnetic Resonance (NMR) results in order to reduce this problem [2,11].

*Part XXI of the series “Applications of Artificial Intelligence in Structure Determination”. For part XX see Ref. [12] in this text.

Fig. 1. Numbering of skeletons (and biogenetic numbering) of diterpenes contained in the database. The numbers in parentheses indicate the numbers of ^{13}C -NMR spectra.

The systems referred to here have not been made to work specifically with natural products whom frequently are substances having more than 15 atoms, this is why the utilization of restrictions in the process of generating structures is fundamental.

In the structural determination of natural products, the phytochemist uses restrictions such as the class of the substance (i.e., flavonoid, diterpene, triterpene, etc.) and the skeleton type. Therefore, if the specialised system tries to simulate the reasoning of the phytochemist it should be able in a first step to recognise the class and the skeleton of the compound.

Fig. 1. (Continued).

Our research group is working on the development of a system named SISTEMAT [5,7,8]. This system contains a program named SISCONST [6] that can analyse the spectra of a compound and provide some restrictions because it recognises substructures and proposes the type of skeleton based on chemical shifts and multiplicity of ^{13}C -NMR signals of the compound. These restrictions are used by the generating program which is presently under development.

^{13}C -NMR spectroscopy is the most widely used technique for structural determination of diterpenes. The compounds in this class of natural products can be grouped into various types of skeletons (Fig. 1). It can be noted that some have more compounds than others, like for example labdanes, clerodanes

Fig. 1. (Continued).

and kauranes (skeletons 61, 76 and 192, respectively). The compounds of such skeletons can be grouped by their functional groups and the presence of heteroatoms in specific positions as long as the carbon atoms corresponding to these positions give characteristic chemical shifts.

Within this work we have verified the way of regrouping the compounds of a specific skeleton and afterward the chemical shift intervals in ^{13}C -NMR that characterises the compounds of a given group. These intervals have been introduced using a consultation program called DITREGRA.

Fig. 1. (Continued).

2. Experimental

For the purpose of this work a database was developed containing the codes and ^{13}C -NMR chemical shifts of 2745 diterpenes distributed between 214 skeletons (Fig. 1). With these data we were able to build a database with the help of input module of the SISTEMAT [4,7].

In order to meet the intervals of chemical shifts that characterise the various groups of substances of a given skeleton, four programs were used: TIPCARB, PICKSIS, SISPICK and PICKRVSF [3].

Fig. 1. (Continued).

Compounds from the skeleton kaurane (Fig. 2) were chosen as an example to demonstrate the process of creation of heuristic rules.

The program TIPCARB provides a table showing the substitution patterns of each atom that belongs to a given type of skeleton. For kauranes (skeleton 192) the results showed by this program (Table 1) that the atoms of any position, except logically quaternary [4,8,10] and carbon 5 (CH), can be oxidised.

The intervals of ^{13}C -NMR chemical shifts of those carbons do not characterise the compounds of these skeletons, since the majority of diterpenes has two or more quaternary carbons with chemical

Fig. 1. (Continued).

shifts in these intervals. Also, the majority of bi-, tri-, and tetra-cyclic diterpenes, have no substitution on carbon 5, therefore indicating that the chemical shift intervals for these types of diterpenes are the same and consequently don't serve to characterise these carbons.

Since these chemical shift intervals for these carbons are not sufficient to determine a diterpene skeleton, it has been proposed to choose carbon atoms that regroup a large number of compounds, even if they represent functional groups. For this, we have chosen quaternary carbons, as well as carbon atoms which are methyl groups on the skeleton without the heterostoms. In the case of compounds with

Fig. 1. (Continued).

a kaurane skeleton, all compounds were pooled within six groups: kaur-16-ene, kaur-16-en-13-OH, kaur-16-OH, kaur-15-ene, kaur-15-en-9-OH and kaurane (Fig. 3).

Following the choice of the carbon atoms to study, we used the programs PICKSIS and SISPICK [3] to obtain the respective chemical shift intervals. For the kaur-16-enes, the intervals found are: 59.2–32.4 (s, C4); 64.3–38.2 (d, C5); 63.2–35.9 (s, C8); 70.0–42.0 (d, C9); 51.5–33.5 (s, C10); 55.5–35.0 (d, C13); 161.6–124.2 (s, C16); 121.1–102.4 (t, C17).

Fig. 1. (Continued).

These intervals were introduced into the program PICKRVSF [3] that was used to do the research based on a comparison with the ^{13}C -NMR spectra of all the diterpenes contained in the database. The list obtained shows that all the compounds that can be kaur-16-ene do not show carbon atoms with chemical shifts and multiplicity in these intervals.

The list obtained shows that 301 compounds distributed into many skeletons and a few groups of kauranes (kaur-15-en-13-OH and kaur-16-ene-9-OH) have chemical shifts and multiplicities in the intervals found for kaur-16-ene (Table 2). These results demonstrate a percentage of recognition equal to 47.80% (144 kaur-16-ene between 301 compounds).

Fig. 1. (Continued).

Kauranes have three quaternary carbon atoms and, as can be observed, some compounds of skeletons 50, 61, 76, 102, 120, 123, 140, 145, 165, 182, 193, 202 and 205 that have a different number of quaternary carbons are confounded. The number of sp^3 quaternary carbons existing in a compound is easily recognised in a ^{13}C -NMR spectra, either by its multiplicity (singlet) or by the intensity of

Fig. 2. Biogenetic numbering of compounds with kaurane skeleton.

Table 1

Results from the program TIPCARB with compounds that present the skeleton kaurane and for which the ^{13}C -NMR spectra are included in the database. Asterisks indicate aromatic carbons and the letter T a triple bond

Atoms	CH_3	CH_2	CH	C	$\text{CH}_2=$	$\text{CH}=$	$\text{C}=$	$\text{HCT}=$	CT	HC^*	C^*	$=\text{C}=$
1	0	242	21	0	0	4	1	0	0	0	0	0
2	0	256	7	0	0	5	0	0	0	0	0	0
3	0	193	54	2	0	0	19	0	0	0	0	0
4	0	0	0	268	0	0	0	0	0	0	0	0
5	0	0	268	0	0	0	0	0	0	0	0	0
6	0	219	43	0	0	2	4	0	0	0	0	0
7	0	155	80	14	0	2	17	0	0	0	0	0
8	0	0	0	268	0	0	0	0	0	0	0	0
9	0	0	248	17	0	0	3	0	0	0	0	0
10	0	0	0	268	0	0	0	0	0	0	0	0
11	0	218	39	0	0	6	5	0	0	0	0	0
12	0	245	19	0	0	3	1	0	0	0	0	0
13	0	0	242	26	0	0	0	0	0	0	0	0
14	0	223	44	0	0	0	1	0	0	0	0	0
15	0	123	56	0	0	25	64	0	0	0	0	0
16	0	0	11	52	0	0	205	0	0	0	0	0
17	41	42	0	0	180	2	3	0	0	0	0	0
18	206	52	0	0	0	2	8	0	0	0	0	0
19	140	29	0	0	0	0	99	0	0	0	0	0
20	232	28	8	0	0	0	0	0	0	0	0	0

the signal. If we consider this fact, and consequently ignore the compounds that present a number of quaternary carbon that is different, the percentage of recognition increases to 61.00% (144 kaur-16-enes between 236 substances with three quaternary carbon atoms).

The number of methyl groups varies between the skeletons. These can be oxidised to alcohol (or ether), aldehyde, acid (or ester) or form a double bond terminal or exo-cyclic. These functions present ^{13}C -NMR chemical shifts that are more or less characteristic. Therefore we can consider that the number of methyl groups of a compound with no functional group is easily deducible from the ^{13}C -NMR spectra. These considerations make up for the kauranes, not to be confounded with the compounds of skeletons 134, 141, 171, 201 and 212. With this reasoning the percentage of recognition becomes 84.70% (144 kaur-16-enes between 170 compounds with the same number of quaternary carbons and same number of methyl groups in the skeletons without the chemical functional groups).

Therefore, it was possible to obtain rules such as: “if the ^{13}C -NMR spectra of the compound under study shows chemical shifts and multiplicity in the intervals 161.6–142.2 (s, C16); 121.1–102.4 (t, C17);

Fig. 3. Groups of compounds belonging to the kaurane skeleton. The asterisks indicate the carbons that were used to create the heuristic rules.

70.0–42.0 (d, C9); 64.3–38.2 (d, C5); 63.2–35.9 (s, C8); 59.2–32.4 (s, C4); 55.5–35.0 (d, C13); 51.5–33.5 (s, C10) therefore there exist a 84.70% probability that the compound is a kaur-16-ene”.

Following the process described for kaur-16-ene, it has been possible to obtain 103 rules for many groups of substances from various skeletons of diterpenes (Table 3). These rules are used as a database for a consultation program which allows the user to verify in the skeleton the group to which the compound under study belongs.

To use the consultation program, the user must introduce the chemical shifts and multiplicities of an unknown. Then the program checks for quaternary carbons using the shift range (70.0–35.0 (s)). The program calculates the number of corresponding signals and verifies which are the diterpene skeletons present in the data base that has the same number of quaternary sp^3 carbons. At this level the program abandons the skeletons which don't match this number.

Following this the program verifies the spectral signals corresponding to methyl groups, by looking at the multiplicity (quartet) of this, and adds the chemical shifts of oxidised atoms ($-CH_2OH$ or $-CH_2OR$, $-CHO$, $-CO_2H$ or CO_2R , $=CH_2$ (t) and CHO). Again the program counts the number of carbon atoms that could be methyl groups in the skeleton without functional groups.

Table 2
Number of compounds separated by skeleton that present a
 ^{13}C -NMR chemical shift proven to be kaurane-16-enes

Skeletons	Number of compounds
kaur-16-en	144
kaur-16-en-13-OH	3
kaur-16-en-9-OH	4
50	11
61	22
76	4
102	1
120	1
123	4
134	41
140	3
141	1
145	1
165	3
171	15
182	5
193	6
201	8
202	2
205	2
212	1
190	18
196	1
<i>total</i>	<i>301</i>

In the last two steps, the user can alter the answer of the computer since sometimes the quaternary atoms don't reflect chemical shifts in the range of interval used by the program. The same can be applied to other functional groups.

This way, the number of skeletons to be searched can be reduced, since the search will be done in one of the 56 sets of skeletons that present the same number of sp^3 quaternary carbon atoms and the same number of methyl groups. After this the spectral signals of the compound under study (chemical shifts and multiplicity) are compared with the chemical shift intervals considered like heuristic rules and the percentage of recognition is presented.

3. Results and discussion

In order to test the efficiency of the program, we have used the ^{13}C -NMR spectral data of 100 diterpenes. For example, we have introduced in the program the ^{13}C -NMR chemical shifts and the signals multiplicity's for the compound in Fig. 4. The program DITREGRA indicates, correctly, that the compound is a kaur-16-ene with a percentage of recognition of 84.70%.

When the group of compound is not present, the program offers skeleton possibilities by comparing the number of quaternary carbon atoms and the number of carbons that could be methylated into the skeleton but without the chemical functional groups. For example, the compound in Fig. 5, when it is without the heterostoms, has 20 carbon atoms, 2 sp^3 quaternary atoms and 7 methyl groups. With these characteristics, the number of possible skeletons for the structure of the compound is reduced from 214 to 7 which corresponds to skeletons 30, 50, 54, 84, 90, 92 and 100.

Table 3

Chemical shifts intervals characteristic of sub-skeletons. (The number before the brackets indicates the numbers of the skeletons described in Fig 1. The prefix EP stands for epoxide)

Intervals of chemical shifts (ppm)	Recognition (%)	Sub-skeleton
168.3–132.5(s,C7); 165.0–122.5(d,C6); 163.6–133.8(s,C3); 155.8–127.5(s,C15); 154.1–120.0(d,C14); 155.1–122.3(d,C10); 143.8–129.6(s,C11); 141.1–115.1(d,C2)	38.6	1[2,6,10,14ENE]
149.3–124.6(d,C6); 142.8–123.6(d,C10); 141.3–130.5(s,C15); 141.0–131.6(s,C11); 139.0–131.5(s,C7); 138.9–134.3(d,C20); 129.1–122.5(d,C14); 125.0–118.0(s,C3)	100.0	1[3(20),6,10,14ENE; 20OR]
85.0–83.6(s,C8); 72.3–71.6(s,C4); 48.7–48.5(d,C1); 32.0–32.0(d,C15)	100.0	10[4OH; 8OR]
154.5–134.0(s,C15); 154.1–124.8(d,C11); 147.5–121.4(d,C7); 144.6–128.6(s,C12); 140.8–131.5(s,C8); 139.3–132.0(s,C4); 133.5–108.6(t,C16); 128.6–121.8(d,C3); 60.4–36.6(d,C1)	73.7	12[3,7,11,15ENE]
159.3–131.6(s,C8); 133.2–126.9(d,C7); 72.6–71.3(s,C4); 62.7–58.5(d,C11); 62.0–59.2(s,C12); 48.0–45.5(d,C1); 33.2–31.5(d,C15)	100.0	12[7ENE; 4OH; 11EP]
151.1–141.3(s,C1); 151.2–122.1(d,C11); 139.5–132.6(s,C10); 136.0–132.6(s,C6); 130.8–126.0(d,C7); 120.8–112.4(t,C19); 71.0–70.6(s,C14); 57.5–49.4(d,C2); 44.0–37.0(d,C3)	100.0	19[1(19),6,10ENE; 14OH]
151.1–146.5(s,C1); 142.6–140.6(d,C17); 141.1–134.3(s,C14); 135.8–133.1(s,C6); 131.1–123.9(d,C7); 119.8–119.0(d,C13); 116.0–113.0(t,C19); 115.9–113.3(s,C10); 49.5–49.2(d,C2); 37.2–36.7(d,C3)	100.0	19[1(19),6,10(17),13ENE; 17OR]
161.1–134.0(s,C4); 142.0–121.5(d,C10); 139.3–129.6(s,C11); 132.6–132.0(s,C15); 127.1–120.1(d,C3); 124.1–123.5(d,C14); 51.2–34.4(d,C1); 44.4–24.1(d,C7)	100.0	21[3,10,14ENE]
158.8–137.3(s,C12); 145.3–142.6(d,C7); 143.8–133.8(s,C2); 135.6–133.3(s,C6); 128.6–122.0(d,C3); 127.3–118.1(d,C13); 35.9–30.2(d,C9); 32.4–26.6(s,C15); 29.2–23.2(d,C8)	100.0	33[2,6,12ENE]
142.6–133.5(s,C5); 120.8–117.4(t,C16); 86.0–83.9(s,C8); 45.5–44.4(d,C17); 45.0–38.2(s,C1); 43.2–36.5(d,C11); 39.4–32.7(d,C10)	100.0	39[5(16)ENE; 8OH]
168.5–138.6(s,C2); 142.4–128.1(s,C1); 132.6–131.5(s,C14); 124.5–123.5(d,C13); 50.2–47.4(d,C3); 32.9–32.3(d,C10); 29.3–28.3(d,C6); 28.7–10.2(d,C7); 26.1–19.5(d,C9)	100.0	41[1,13ENE]

Table 3
(Continued)

Intervals of chemical shifts (ppm)	Recognition (%)	Sub-skeleton
161.3–149.6(s,C10); 141.5–140.8(s,C4); 131.6–131.3(s,C15); 124.9–124.5(d,C14); 124.0–123.8(d,C3); 107.1–104.0(t,C18); 61.3–60.5(d,C5); 47.7–43.7(d,C7); 46.2–43.0(d,C1); 35.0–34.7(d,C11)	100.0	46[3,10(18),14ENE]
154.1–152.5(s,C10); 115.5–114.7(t,C20); 57.9–54.7(d,C9); 55.5–54.2(d,C5); 54.8–36.3(d,C14); 54.2–32.0(d,C13); 48.8–36.1(s,C8); 48.4–30.1(s,C4)	100.0	50[10(20)ENE]
142.0–129.1(s,C10); 138.8–122.3(s,C9); 137.8–130.8(s,C15); 134.3–121.1(s,C6); 124.8–124.0(d,C14); 48.7–31.5(d,C11); 47.2–32.9(d,C4); 41.5–26.5(d,C1)	83.3	51[5,7,9,14ENE]
167.8–140.1(s,C9); 132.5–125.0(s,C8); 52.0–45.4(d,C5); 41.2–37.0(s,C10); 40.7–33.2(s,C4); 31.4–30.7(d,C13)	100.0	61[8ENE]
166.7–140.1(s,C9); 143.3–138.8(d,C16); 132.6–126.5(s,C8); 126.0–123.5(s,C13); 52.2–39.0(s,C4); 51.4–45.6(d,C5); 39.5–36.2(s,C10)	100.0	61[8,13(16)ENE; 16OR]
166.1–137.8(s,C9); 161.5–133.6(s,C13); 146.5–115.4(d,C14); 131.7–122.3(s,C8); 51.9–45.7(d,C5); 42.3–33.0(s,C4); 42.0–36.9(s,C10)	100.0	61[8,13ENE]
143.2–138.3(d,C16); 127.1–124.2(s,C13); 82.9–75.1(s,C9); 59.7–44.7(d,C5); 53.2–33.5(s,C4); 51.4–31.1(d,C8); 47.7–38.6(s,C10)	71.5	61[13(16)ENE; 9OH; 16OR]
142.6–134.8(s,C13); 123.6–116.5(d,C14); 88.6–79.3(d,C12); 81.6–80.0(s,C8); 65.7–54.1(d,C9); 65.7–52.3(d,C5); 54.1–36.2(s,C10); 34.0–32.0(s,C4)	100.0	61[13ENE; 8(12)EP]
85.8–81.5(d,C12); 81.4–80.8(s,C8); 74.5–72.5(s,C13); 61.1–60.0(d,C9); 57.3–57.0(d,C5); 36.5–36.2(s,C10); 33.2–33.0(s,C4)	100.0	61[13OH; 8(12)EP]
150.3–144.8(s,C13); 115.5–112.6(t,C16); 75.5–72.5(s,C8); 62.4–54.4(d,C9); 61.5–49.2(d,C5); 49.7–33.2(s,C4); 40.0–38.5(s,C10)	93.3	61[13(16)ENE; 8OH]
164.8–134.4(s,C13); 136.5–114.9(d,C14); 75.5–71.9(s,C8); 65.1–53.4(d,C9); 60.9–47.0(d,C5); 48.5–32.7(s,C4); 40.0–37.7(s,C10)	70.3	61[13ENE; 8OH]
75.0–72.3(s,C13); 74.8–73.0(s,C8); 65.9–61.4(d,C9); 56.2–55.8(d,C5); 39.3–37.5(s,C10); 33.4–33.0(s,C4)	100.0	61[8,13OH]

Table 3
(Continued)

Intervals of chemical shifts (ppm)	Recognition (%)	Sub-skeleton
149.6–144.5(s,C8); 112.0–106.8(t,C17); 57.0–51.0(d,C9); 56.9–46.9(d,C5); 47.7–33.0(s,C4); 39.7–35.7(s,C10); 31.1–28.7(d,C13)	100.0	61[8(17)ENE]
150.1–134.3(s,C8); 139.3–121.6(d,C7); 63.9–50.9(d,C9); 56.7–43.7(d,C5); 47.2–32.2(s,C4); 42.5–35.4(s,C10); 31.3–30.2(d,C13)	100.0	61[7ENE]
164.1–138.6(s,C13); 138.9–133.6(s,C8); 129.0–121.6(d,C7); 127.0–115.1(d,C14); 54.9–44.5(d,C5); 54.7–49.4(d,C9); 51.0–34.9(s,C10); 42.1–32.2(s,C4)	100.0	61[7,13ENE]
159.8–131.3(d,C12); 150.1–146.1(s,C8); 140.1–123.3(s,C13); 109.9–104.1(t,C17); 62.2–49.9(d,C9); 57.5–48.4(d,C5); 47.8–33.5(s,C4); 40.2–38.4(s,C10)	67.9	61[8(17),12ENE]
138.8–135.1(s,C8); 126.5–122.0(d,C7); 73.5–72.8(s,C13); 55.2–49.9(d,C9); 55.2–42.9(d,C5); 47.7–37.2(s,C4); 43.2–36.5(s,C10)	100.0	61[7ENE; 13OH]
149.5–145.3(s,C8); 113.3–106.2(t,C17); 76.1–73.0(s,C13); 57.4–48.7(d,C9); 56.0–40.7(d,C5); 47.7–33.0(s,C4); 43.7–39.2(s,C10)	100.0	61[8(17)ENE; 13OH]
152.2–145.1(s,C8); 146.8–138.6(d,C16); 128.1–124.4(s,C13); 110.3–104.9(t,C17); 61.5–44.9(d,C9); 56.2–46.2(d,C5); 54.5–33.5(s,C4); 53.0–37.9(s,C10)	94.4	61[8(17),13(16)ENE; 16OR]
164.6–133.3(s,C13); 150.6–144.0(s,C8); 145.3–114.6(d,C14); 113.8–106.1(t,C17); 58.0–26.2(d,C5); 56.9–49.5(d,C9); 55.4–32.7(s,C4); 53.2–38.2(s,C10)	77.2	61[8(17),13ENE]
78.2–74.9(s,C13); 77.1–74.5(d,C14); 76.1–75.1(s,C8); 65.4–62.5(t,C15); 57.2–55.4(d,C5); 55.5–49.0(d,C9); 38.4–36.3(s,C10); 37.7–33.3(s,C4)	100.0	62[14,15OH]
148.3–140.1(d,C14); 115.5–109.5(t,C15); 83.5–73.0(s,C13); 80.6–74.4(s,C8); 70.5–45.2(d,C9); 57.2–43.0(d,C5); 53.8–32.5(s,C4); 42.9–35.9(s,C10)	91.8	62[14ENE]
148.1–144.5(s,C8); 118.0–108.3(t,C17); 93.5–88.6(s,C9); 82.5–81.1(s,C13); 58.5–40.4(d,C5); 43.7–40.7(s,C10); 33.7–32.2(s,C4)	100.0	63[8(17)ENE]
154.3–133.6(s,C8); 137.5–125.6(d,C7); 92.0–69.6(s,C9); 89.9–81.1(s,C13); 56.5–36.7(d,C5); 46.7–32.4(s,C4); 45.9–40.2(s,C10)	86.7	63[7ENE]

Table 3
(Continued)

Intervals of chemical shifts (ppm)	Recognition (%)	Sub-skeleton
76.5–71.7(s,C8); 65.9–57.5(d,C9); 57.4–53.5(d,C5); 39.2–37.3(s,C10); 34.2–33.1(s,C4)	100.0	68[8OH]
166.0–143.1(s,C13); 141.3–140.8(s,C10); 128.1–114.6(d,C14); 120.5–119.6(d,C1); 45.0–44.7(s,C4); 43.5–42.7(s,C9); 38.9–38.5(d,C8); 38.5–37.9(d,C5)	100.0	72[1(10),13ENE]
175.8–168.4(s,C13); 116.2–113.7(d,C14); 67.4–60.7(s,C4); 57.6–45.5(d,C10); 55.0–46.9(t,C18); 47.0–45.0(s,C5); 39.7–38.0(s,C9); 39.6–32.5(d,C8)	100.0	76[13ENE; 4(18)EP]
66.5–60.5(s,C4); 51.7–41.8(d,C10); 49.5–42.0(t,C18); 46.3–40.0(d,C13); 46.1–44.7(s,C5); 41.6–39.7(s,C9); 36.1–32.7(d,C8)	100.0	76[4(18)EP]
162.6–140.9(s,C4); 156.6–120.1(d,C3); 136.6–135.5(s,C13); 129.1–126.5(d,C12); 55.9–49.2(s,C5); 44.7–36.5(d,C10); 39.2–37.5(s,C9); 36.8–35.0(d,C8)	100.0	76[3,12ENE]
148.1–138.1(d,C16); 131.1–124.0(s,C13); 68.1–55.7(s,C4); 57.5–39.5(d,C10); 56.8–39.9(s,C5); 56.5–37.7(s,C9); 54.2–42.0(t,C18); 51.0–39.9(s,C8)	100.0	76[13(16)ENE; 16OR; 4(18)EP]
171.0–136.1(s,C4); 136.8–120.4(d,C3); 51.4–36.0(d,C8); 48.2–41.4(d,C10); 48.0–37.6(s,C5); 43.5–38.0(s,C9); 35.9–26.0(d,C13)	100.0	76[3ENE]
173.1–169.8(s,C13); 143.8–114.0(d,C14); 70.5–60.5(s,C4); 63.7–57.2(d,C3); 53.4–37.7(s,C5); 45.5–36.5(d,C10); 38.5–37.2(s,C9); 36.9–30.1(d,C8)	100.0	76[13ENE; 3EP]
139.5–138.3(d,C16); 125.5–124.8(s,C13); 65.6–61.9(s,C4); 60.5–57.4(d,C3); 53.5–38.2(s,C9); 51.0–39.9(d,C10); 47.0–36.5(s,C5); 40.9–30.8(d,C8)	100.0	76[13(16)ENE; 16OR; 3EP]
174.0–132.1(s,C13); 172.1–134.8(s,C4); 146.8–112.9(d,C14); 141.6–118.3(d,C3); 54.7–30.6(d,C8); 51.4–37.7(s,C5); 48.4–37.4(d,C10); 48.0–36.5(s,C9)	58.1	76[3,13ENE]
140.8–139.0(d,C16); 126.0–122.5(s,C13); 68.2–37.0(d,C4); 64.0–35.0(d,C10); 59.9–41.4(s,C5); 52.5–33.0(d,C8); 51.2–34.5(s,C9)	25.0	76[13(16)ENE; 16OR]
147.3–138.6(d,C16); 132.1–123.8(s,C13); 82.9–74.4(s,C4); 56.4–34.2(s,C9); 53.9–37.2(d,C10); 53.4–34.0(s,C5); 50.5–33.0(d,C8)	70.0	76[13(16)ENE; 4OH; 16OR]

Table 3
(Continued)

Intervals of chemical shifts (ppm)	Recognition (%)	Sub-skeleton
174.0–161.3(s,C13); 115.9–114.4(d,C14); 85.0–60.7(s,C4); 56.7–42.4(s,C5); 50.3–40.2(d,C10); 39.9–36.7(s,C9); 36.7–31.1(d,C8)	57.1	76[13ENE; 4OH]
173.8–170.1(s,C13); 115.5–115.3(d,C14); 56.0–42.0(d,C4); 53.0–40.0(d,C10); 50.0–43.0(s,C5); 39.4–38.2(s,C9); 37.2–30.1(d,C8)	100.0	76[13ENE]
169.2–127.4(s,C4); 144.3–116.1(d,C3); 144.2–136.0(d,C16); 139.0–122.6(s,C13); 59.1–36.0(s,C5); 59.1–34.2(s,C9); 53.7–30.3(d,C8); 52.3–32.7(d,C10)	66.2	76[3,13(16)ENE; 16OR]
163.6–141.9(s,C4); 147.1–144.3(s,C13); 128.7–120.6(d,C3); 116.0–115.1(t,C16); 54.7–35.0(d,C8); 53.5–37.7(s,C5); 47.2–35.2(d,C10); 38.9–37.4(s,C9)	100.0	76[3,13(16)ENE]
167.1–128.6(s,C5); 145.0–139.5(d,C16); 144.5–123.0(s,C4); 126.5–123.1(s,C13); 56.9–38.2(s,C9); 49.7–33.0(d,C8); 43.1–33.2(d,C10)	86.7	77[4,13(16)ENE; 16OR]
172.5–162.6(s,C9); 166.8–158.8(s,C14); 112.5–105.9(s,C8); 108.0–105.0(d,C11); 54.7–46.7(d,C5); 50.0–42.2(s,C4); 42.5–35.7(s,C10); 41.2–29.5(d,C15)	100.0	85[8(14),9(11)ENE]
162.5–154.0(s,C9); 143.0–127.3(s,C8); 132.3–121.0(d,C7); 115.4–109.8(d,C11); 52.5–45.2(d,C5); 49.5–42.7(s,C4); 42.4–29.5(d,C15); 38.2–34.5(s,C10)	100.0	85[7,9(11)ENE]
159.6–156.8(s,C9); 118.9–116.6(d,C11); 58.9–57.4(s,C8); 54.9–53.5(d,C7); 52.7–42.0(s,C4); 45.4–41.7(d,C5); 36.5–35.4(s,C10); 34.7–24.1(d,C15)	100.0	85[9(11)ENE; 7EP]
148.0–130.6(s,C13); 145.3–132.0(s,C9); 46.4–40.2(d,C5); 43.0–34.5(s,C10); 42.5–34.2(s,C4); 36.2–26.6(d,C15)	83.3	93[8,11,13ENE]
146.1–122.4(d,C3); 142.6–124.9(d,C14); 136.1–129.8(s,C4); 134.8–127.5(s,C15); 56.2–50.0(d,C10); 52.0–40.0(d,C7); 49.7–39.4(s,C1); 33.7–33.0(d,C11)	100.0	103[3,14ENE]
140.4–136.8(s,C6); 117.6–116.4(d,C5); 75.7–73.3(s,C4); 72.8–71.0(s,C15); 43.3–42.9(d,C13); 31.9–29.1(d,C2); 31.0–30.4(d,C11); 28.9–24.7(d,C9); 24.7–19.0(s,C10)	100.0	110[5ENE; 4(15)EP]
153.9–136.5(s,C12); 147.9–123.2(d,C13); 143.3–136.7(s,C3); 141.6–135.5(s,C4); 55.7–48.1(s,C9); 54.8–41.0(s,C6); 40.7–38.5(d,C5); 27.2–25.7(d,C18)	100.0	120[3,12ENE]

Table 3
(Continued)

Intervals of chemical shifts (ppm)	Recognition (%)	Sub-skeleton
84.1–56.0(s,C8); 56.0–50.0(d,C9); 55.4–45.2(d,C5); 50.5–36.2(s,C10); 47.5–33.2(s,C4); 46.5–30.5(s,C13)	63.6	134[8OR]
61.3–32.1(s,C13); 56.2–43.1(d,C9); 49.7–41.0(d,C5); 47.8–36.0(s,C4); 45.6–35.1(d,C8); 39.0–35.6(s,C10)	96.4	134[without substitution]
76.9–70.4(s,C8); 59.3–51.2(d,C9); 57.7–42.1(d,C5); 47.4–33.0(s,C4); 42.6–35.0(s,C13); 41.6–35.7(s,C10)	100.0	134[8OH]
177.3–145.1(s,C9); 121.9–113.0(d,C11); 54.2–40.2(d,C5); 47.4–33.5(s,C4); 46.4–31.3(s,C13); 39.9–37.9(s,C10); 39.2–28.7(d,C8)	86.7	134[9(11)ENE]
144.5–121.9(d,C14); 140.6–134.3(s,C8); 59.2–46.0(d,C9); 57.0–32.5(s,C13); 56.2–39.5(d,C5); 47.7–33.0(s,C4); 47.4–35.7(s,C10)	77.7	134[8(14)ENE]
166.6–134.4(s,C9); 153.6–122.6(s,C8); 54.2–33.9(d,C5); 47.4–32.7(s,C4); 46.0–31.0(s,C13); 43.5–36.9(s,C10)	75.6	134[8ENE]
148.0–133.0(s,C8); 134.8–114.0(d,C7); 55.7–44.7(d,C9); 53.8–32.5(s,C4); 52.7–37.2(d,C5); 49.7–35.4(s,C13); 42.2–30.8(s,C10)	54.9	134[7ENE]
158.1–149.8(s,C9); 144.2–131.3(s,C8); 48.9–47.4(d,C5); 38.2–36.5(d,C10); 37.8–35.2(s,C13); 32.9–32.7(s,C4)	100.0	135[8ENE]
146.6–138.1(s,C5); 123.3–115.5(d,C6); 47.7–36.9(d,C10); 45.5–34.6(s,C4); 40.5–34.5(s,C9); 37.5–35.7(s,C13); 36.8–34.7(d,C8)	100.0	137[5ENE]
139.3–135.6(s,C10); 135.7–125.0(s,C5); 53.7–36.5(s,C9); 47.7–33.9(s,C4); 45.0–37.2(d,C8); 42.4–36.5(s,C13)	100.0	137[5(10)ENE]
152.8–150.5(s,C13); 107.7–104.3(t,C17); 55.7–52.4(d,C5); 55.0–46.9(d,C9); 49.5–49.0(d,C14); 46.7–40.7(d,C8); 3.8–32.0(s,C4); 39.5–37.0(s,C10)	100.0	139[13(17)ENE]
155.1–138.8(s,C9); 146.6–129.1(s,C14); 137.1–120.8(s,C13); 134.8–123.8(s,C8); 55.4–47.4(d,C5); 53.5–37.0(s,C10); 47.5–33.2(s,C4)	51.4	139[8,11,13ENE]
61.4–37.8(d,C9); 60.5–41.5(d,C14); 58.9–47.4(d,C5); 55.5–28.3(d,C13); 52.0–34.8(s,C8); 39.7–37.3(s,C10); 33.7–32.7(s,C4)	95.2	141[without substitution]

Table 3
(Continued)

Intervals of chemical shifts (ppm)	Recognition (%)	Sub-skeleton
154.1–120.8(d,C12); 137.8–126.8(s,C13); 62.5–51.2(d,C14); 59.2–51.0(d,C9); 57.4–56.0(d,C5); 43.5–33.0(s,C4); 38.5–37.1(s,C10); 37.0–34.2(s,C8)	100.0	141[12ENE]
154.0–139.5(s,C9); 151.1–130.3(s,C13); 147.3–127.4(s,C8); 76.8–71.5(s,C15); 52.8–43.2(d,C5); 44.0–33.2(s,C4); 39.3–37.1(s,C10)	100.0	145[8,11,13ENE; 15OH]
155.5–145.1(s,C13); 130.3–125.1(s,C15); 65.9–58.7(s,C8); 56.2–54.4(d,C14); 55.9–47.9(d,C5); 53.0–33.5(s,C4); 51.9–38.7(d,C9); 41.5–35.5(s,C10)	100.0	145[13(15)ENE; 8(14)EP]
156.6–150.6(s,C12); 155.1–139.3(s,C9); 149.8–132.8(s,C8); 126.0–119.4(s,C13); 52.2–40.0(d,C5); 41.5–33.0(s,C10); 39.2–32.5(s,C4); 30.0–22.8(d,C15)	92.3	145[8,12ENE; 12OH]
156.5–123.0(s,C9); 148.7–117.0(s,C13); 141.8–105.0(s,C8); 61.5–42.0(d,C5); 52.0–32.0(s,C4); 51.7–36.2(s,C10); 36.5–23.1(d,C15)	78.2	145[8,11,13ENE]
148.0–133.3(s,C9); 145.5–130.0(s,C14); 134.1–124.0(s,C8); 52.2–49.3(d,C5); 44.7–30.3(s,C4); 38.5–37.8(s,C10); 27.2–26.6(d,C15)	100.0	157[8,11,13ENE]
51.2–46.5(s,C4); 48.5–38.7(d,C9); 45.2–40.6(d,C5); 42.1–24.5(d,C8); 39.2–32.2(s,C10)	100.0	162[without substitution]
155.0–147.3(s,C9); 135.4–123.8(s,C8); 52.9–48.9(d,C5); 48.6–31.9(s,C4); 38.6–33.2(s,C10)	100.0	162[8,11,13ENE]
161.6–124.2(s,C16); 121.1–102.4(t,C17); 70.0–42.0(d,C9); 64.3–38.2(d,C5); 63.2–35.9(s,C8); 59.2–32.4(s,C4); 55.5–35.0(d,C13); 51.5–33.5(s,C10)	93.7	171[16ENE]
75.6–72.5(s,C11); 72.0–71.6(s,C4); 60.2–53.9(d,C5); 47.4–45.2(s,C9); 42.0–32.5(d,C6); 37.0–36.5(s,C15)	100.0	175[4,11OH]
87.9–79.2(s,C5); 86.9–78.2(s,C16); 78.5–77.2(s,C10); 59.0–50.8(d,C13); 58.5–47.5(d,C1); 57.0–46.4(s,C4); 56.8–46.5(s,C8); 56.0–47.2(d,C9)	100.0	182[5,10,16OH]
153.0–147.1(s,C10); 112.6–110.4(t,C20); 87.5–83.5(s,C5); 81.5–79.5(s,C16); 54.7–47.9(d,C13); 53.7–50.4(s,C4); 52.3–47.2(d,C9); 50.6–44.7(s,C8); 47.7–44.5(d,C1)	100.0	182[10(20)ENE; 5,16OH]

Table 3
(Continued)

Intervals of chemical shifts (ppm)	Recognition (%)	Sub-skeleton
137.8–133.8(s,C6); 136.1–133.4(s,C2); 132.3–131.5(d,C1); 131.8–123.1(d,C7); 86.0–84.9(s,C4); 72.3–71.8(s,C10); 43.8–43.0(d,C8); 38.7–38.5(d,C11); 28.5–24.0(s,C15); 24.1–23.1(d,C13); 24.0–23.2(d,C14)	100.0	185[1,6ENE; 4OR]
162.0–147.6(s,C1); 126.3–121.1(d,C2); 79.1–72.0(s,C8); 61.0–54.5(s,C4); 58.5–51.3(d,C7); 49.3–47.2(s,C9); 46.7–38.2(s,C6); 35.5–28.9(d,C15)	100.0	186[1ENE]
74.0–66.6(s,C8); 61.7–53.6(d,C1); 61.2–51.9(s,C4); 60.6–47.9(d,C7); 52.9–47.9(s,C9); 47.8–38.3(s,C6); 33.3–27.9(d,C15)	100.0	186[without substitution]
138.6–137.0(s,C13); 128.3–123.3(d,C14); 51.7–50.7(s,C9); 50.7–41.7(d,C5); 44.2–36.2(s,C4); 42.5–42.2(d,C8); 41.7–37.9(d,C12); 38.4–36.4(s,C10)	100.0	188[13ENE]
153.1–141.6(s,C16); 115.7–104.3(t,C17); 58.4–49.0(d,C5); 55.2–39.0(d,C9); 53.2–33.3(s,C8); 51.9–38.6(s,C4); 45.0–35.5(d,C12); 39.7–37.4(s,C10)	81.8	190[16ENE]
82.8–71.9(s,C16); 60.2–48.9(d,C9); 59.0–49.9(d,C5); 50.0–32.0(d,C12); 47.6–33.0(s,C4); 47.2–32.7(s,C8); 42.0–32.5(s,C10)	36.9	190[16OH]
158.5–127.1(d,C15); 149.3–135.8(s,C16); 62.5–47.9(s,C8); 56.9–41.5(d,C5); 56.0–43.0(d,C9); 53.7–36.7(d,C13); 52.4–33.2(s,C4); 40.5–33.7(s,C10)	87.5	192[15ENE]
158.0–149.8(s,C16); 114.1–102.8(t,C17); 84.5–76.9(s,C9); 57.5–48.5(s,C8); 55.2–38.4(d,C5); 47.9–33.2(s,C4); 47.5–43.2(s,C10); 43.2–37.5(d,C13)	100.0	192[16ENE; 9OH]
84.3–73.3(s,C16); 62.4–49.2(d,C9); 61.2–39.5(d,C13); 57.5–48.2(d,C5); 47.4–33.2(s,C4); 46.4–38.6(s,C8); 42.0–33.1(s,C10)	74.5	192[16OH]
63.6–41.9(d,C9); 61.2–45.0(s,C8); 57.4–43.2(d,C16); 57.2–32.7(d,C13); 57.0–42.7(d,C5); 44.0–33.0(s,C4); 41.5–35.9(s,C10)	10.2	192[without substitution]
163.1–154.1(s,C16); 112.6–102.9(t,C17); 81.9–74.9(s,C13); 59.7–41.5(s,C8); 59.2–47.0(d,C9); 59.2–34.7(s,C4); 58.2–47.7(d,C5); 41.4–34.2(s,C10)	100.0	192[16ENE; 13OH]

Table 3
(Continued)

Intervals of chemical shifts (ppm)	Recognition (%)	Sub-skeleton
161.6–124.2(s,C16); 121.1–102.4(t,C17); 70.0–42.0(d,C9); 64.3–38.2(d,C5); 63.2–35.9(s,C8); 59.2–32.4(s,C4); 55.5–35.0(d,C13); 51.5–33.5(s,C10)	84.7	192[16ENE]
161.0–148.9(s,C16); 115.1–107.1(t,C17); 54.7–50.2(d,C9); 51.7–45.5(s,C8); 49.9–44.8(d,C5); 46.0–37.9(d,C4); 43.5–33.7(s,C10); 43.2–37.6(d,C13)	100.0	193[16ENE]
56.7–54.0(d,C9); 56.7–47.9(d,C5); 49.5–42.8(d,C13); 47.0–33.2(s,C4); 42.8–42.2(s,C8); 39.7–34.3(s,C10)	100.0	194[without substitution]
60.2–44.6(d,C9); 57.5–35.7(s,C13); 57.2–42.2(d,C5); 55.7–44.0(s,C8); 49.9–32.5(s,C4); 44.8–33.2(s,C10)	85.4	197[without substitution]
139.3–132.3(d,C15); 138.0–129.8(d,C16); 57.5–43.5(s,C13); 56.9–42.2(d,C5); 56.0–44.6(d,C9); 55.7–44.0(s,C8); 49.9–32.5(s,C4); 44.8–33.2(s,C10)	100.0	197[15ENE]
158.2–153.0(s,C16); 109.5–107.1(t,C17); 96.2–91.4(s,C10); 61.0–52.1(d,C9); 58.7–50.2(d,C5); 55.6–46.3(s,C4); 53.0–50.7(d,C6); 52.7–51.3(s,C8); 52.6–39.1(d,C13)	100.0	202[16ENE; 10OR]
65.5–36.9(d,C9); 57.0–44.7(d,C5); 53.2–39.7(s,C8); 53.2–33.0(s,C4); 40.2–19.3(d,C12); 39.7–32.2(s,C10); 31.2–20.1(s,C16); 31.1–21.0(d,C13)	100.0	210[without substitution]

Fig. 4. Structure of a diterpene isolated from *Oxidia angusta* [13].

The program gives the correct group in 64% of cases analysed and includes the answer “non studied” when there exist no chemical shift intervals characteristic of the group. With these answers we have verified that 27 times the skeleton had not been studied, 16 times the correct groups appeared as the answer and unique proposal, 15 times as a first answer and 3 times as a second or third answer. In cases where more than one answer is given, this indicates that the ^{13}C -NMR signal of the compound under study can be present within more than one interval of chemicals characterising the group.

Fig. 5. Structure of a diterpene isolated from *Haplopappus parvifolius* [14].

4. Conclusion

SISTEMAT is until now the only system able to incorporate information on chemical classes, skeletons and botanical data able to operate with large restrictions for any program able to generate structures. In this work we have demonstrated that the group of compounds from the same skeleton can be characterised by its chemical shifts intervals.

DITREGRA presents difficulties in proposing the number of possible methyl groups in the skeleton because the intervals used in the data base, for the functional groups, are very broad. This is why the user must study the spectra to confirm or correct the number of methyl groups proposed by the program. Nevertheless, this error from the program is presently being minimised with the introduction of infra-red and ^1H -NMR data into SISTEMAT.

References

- [1] M. Carabedian, I. Dagane and E. Dubois, *Analytical Chemistry* **60** (1998), 2186.
- [2] B.D. Christie and M.J. Munk, *J. Am. Chem. Soc.* **113** (1991), 3750.
- [3] V.P. Emerenciano, A.C. Bussoline, M. Furlan, G.V. Rodrigues and D.L.G. Fromanteau, *Spectroscopy* **11** (1993), 95.
- [4] V.P. Emerenciano, G.V. Rodrigues and J.P. Gastmans, *Química Nova* **16** (1993), 431.
- [5] 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, *Spectroscopy* **12** (1994), 91.
- [6] D.L.G. Fromanteau, J.P. Gastmans, S.A. Vestri, V.P. Emerenciano and J.H.G. Borges, *Computer & Chemistry* **17** (1993), 369.
- [7] J.P. Gastmans, M. Furlan, M.N. Lopes, J.H.G. Borges and V.P. Emerenciano, *Química Nova* **13** (1990), 10.
- [8] J.P. Gastmans, M. Furlan, M.N. Lopes, J.H.G. Borges and V.P. Emerenciano, *Química Nova* **13** (1990), 75.
- [9] R. Lindsay, B.G. Buchanan, E.A. Feigenbaum and J. Lederberg, *Applications of Artificial Intelligence for Organic Chemistry – The DENDRAL Project*, McGraw-Hill, USA, 1980.
- [10] M.E. Munk, M. Farkas, A.H. Lipkis and B.D. Christie, *Mikrochimica Acta* **II** (1986), 199.
- [11] C. Peng, S. Yuan, C. Zheng and Y.J. Hui, *J. Chem. Inf. Comput. Sci.* **34** (1994), 805.
- [12] G.V. Rodrigues, I.P.A. Campos and V.P. Emerenciano, *Spectroscopy* (1997, in press).
- [13] C. Zdero, F. Bohmann and A. Anderberg, *Phytochemistry* **30** (1991), 2703.
- [14] C. Zdero, F. Bohmann and H.M. Niemeyer, *Phytochemistry* **30** (1991), 3683.