NPC Natural Product Communications

2006
Vol. 1
No. 1
57 - 64

Automatic Structure Elucidation through Data Base Search and 2D NMR Spectral Analysis

Jean-Marc Nuzillard*[a] and Vicente de Paulo Emerenciano[b]

[a]*FRE 2715, University of Reims, Moulin de la Housse, BP 1039, 51687 REIMS Cedex 2, France*

[b]*Instituto de Química, Universidade de São Paulo, Caixa Postal 26077, 05513-970 São Paulo, Brazil*

jm.nuzillard@univ-reims.fr

Received: November 10th, 2005; Accepted: December 19th, 2005

This work shows how two expert systems, LSD and SISTEMAT, can be used together to solve structure elucidation problems that were selected from recent literature articles. The LSD system is a structure generator that mainly relies on homo- and heteronuclear 2D NMR data. It lacks the knowledge of chemical shift values and of natural product chemistry. Conversely, the SISTEMAT data base contains about 20000 natural compounds and refers to both their ^{13}C NMR chemical shifts and their botanical origin. When exploited by dedicated computer programs it yields structural constraints such as skeletal types and ring systems. The botanical and spectroscopic data in SISTEMAT proved to be very complementary in the constraints extraction process. Several application examples of the proposed methodology are described in detail.

Keywords: Structure elucidation, expert systems, chemical databases, LSD, SISTEMAT, 2D NMR.

Since the highly cited and pioneering DENDRAL project, four decades ago, several research groups gave special attention to computer-assisted structural elucidation (CASE) methods. The new methodologies are mainly tested on examples drawn from natural products chemistry literature. This research field provides the most difficult problems, due to the extreme diversity of the possible structures. The DENDRAL project resulted in a collection of about one hundred of papers and some test cases of automatized structure problems are described [1]. The DENDRAL paradigm, "Plan-Generate-Test", shows how the control of the intermediate solutions found during the first steps of molecular fragment assembly can be adjusted by the chemist in order to avoid any combinatorial explosion during the generation phase.

Working with a different approach, Dubois *et al.* elaborated the DARC system [2]. The related EPIOS methodology, or Elucidation by Progressive Intersection of Ordered Substructures [3], is a clever strategy towards simultaneous spectra interpretation and structure generation. However, the publications show that the structure of molecules containing 20 to 30 carbon atoms cannot be elucidated by this system. The EPIOS methodology provides an efficient way to fragment assembly, but working with DEPT spectra only is not always sufficient to avoid a combinatorial explosion.

The group of Munk also developed a CASE system, the efficiency of which was demonstrated in several papers. Working with 1D NMR spectroscopy, they published the structure elucidation of molecules of non-trivial complexity [4]. Their system is now able to accept 2D NMR data, and some complex problems are listed [5].

At BASF, a system based on substructure search within a large database was developed [6]. The corresponding Specsolv software is now a commercial product [7].

In Japan, another important CASE project, named CHEMICS, has been developed since the middle of the 1970' [8].

Presently, the main CASE systems incorporate data from 2D NMR spectroscopy [9-13]. Their features

are reported in two review articles by Jaspers [14] and Steinbeck [15]. Among the cited examples, the structure elucidation of molecules such as ascomycin and polycarpol demonstrate the power of the most recent CASE computer programs. A collection of various compounds from the literature and their computer-assisted elucidation was published recently by Elyashberg [12].

CASE systems that mainly rely on chemical shift correlation data propose structures that need to be sorted according to their chemical and spectroscopic likelihood. In a classical approach, the set of solution structures that is obtained for a given problem can be converted into the corresponding set of predicted properties that have subsequently to be matched against the experimental data. Generally, ^{13}C NMR chemical shift value prediction is chosen as the most adequate strategy. Conversely, the goal of the present work is to select from a solution set only those solutions that contain one or more substructures. The latter are selected in a natural product data base from chemical shift values and/or botanical data. In the present case, the LSD input files are supplemented with constraints that are provided by the SISTEMAT expert system.

The wide availability of reliable heteronuclear chemical shift correlations in the early 1990' has given a very strong impulse to the field of natural product structure elucidation. It became obvious that a thorough knowledge of chemical shift rules was not necessary to find a structure when enough 2D NMR data was available. Following some kind of biomimetic way of problem solving, the LSD software was elaborated and proved to be useful in a wide range of situations [16-18]. A LSD data set contains a list of skeletal atom status (chemical element, hybridization state, number of attached hydrogen atoms, initial bonds (if any), COSY, HSQC and HMBC correlations, atom properties (as constraints on possible neighboring atoms) and substructure description (if any). If all atoms of a substructure are identified in the atom set, the corresponding bonds can be given as the beginning of the solution. Otherwise, the substructure acts as a filtering tool for the generated structures. Complex search operations involving logical operations between the results of many substructure searches can be achieved through a dedicated program that uses LSD in the background.

Structure solving by LSD is a four stage process. During the first one, the data file is analyzed and the structure is initialized by the user-supplied bonds and COSY correlations, if any. The second stage is the repetition of two actions, the selection of an HMBC correlation and the setting of bonds between atoms so that this correlation becomes assigned. The correlation selection order has a strong impact on resolution times and a heuristic criterion based on the author's own experience was implemented: the structure under construction must be kept as connected as possible and the correlations of the atoms that are the closest to have all their neighbors must be considered first. The third stage is the combinatorial pairing of atoms that still have free valences at the end of the second stage. Finally, the double bonds are placed, structures that do not obey Bredt's rule are discarded and the substructure element, if any, if checked for its presence or absence, depending on user's wish.

The concepts were initially developed in PROLOG language and were finally coded in C. LSD can be downloaded from www.univ-reims.fr/LSD as free software.

The expert system named SISTEMAT was elaborated by a group of chemists in Brazil during the early 1990' as a way to specifically assist the resolution of natural products chemistry problems. SISTEMAT deals with ^{1}H and ^{13}C NMR data, mass spectra and botanical data. It can handle up to 20 data types such as UV, IR, pharmacological activity, etc…

The main database contains references to compounds that are then used as indexes in all the other data tables (spectroscopic, botanical or pharmacological). The system was involved in chemotaxonomical studies, such as the occurrence of terpenoids in the Lamiaceae [19] and Asteraceae [20] families. A program named SISTAX was developed to work with botanical data linking a compound structure to taxonomic characters, to export the compound structures and botanical data to a spreadsheet program and artificial neural network systems [21].

In the early development stage, all efforts were made to find a simple codification method that would enable the creation of a complex database using personal computers with only 10 to 20 megabytes of hard disk storage. The retained compound coding scheme was published in 1989 [22]. The compressed storage method made it possible to create a database

that contains 50000 compounds in a personal computer with little available hard disk space. The construction of the full database was performed by collecting smaller ones, each being dedicated to a specific chemical class (monoterpenes, sesquiterpenes, iridoids). Compound records include carbon skeletal types that can be used as constraints in a CASE process. The database now contains about 20000 compounds and about 800 skeletal types. As a single compound can be re-isolated many times, any new report is linked to the list of botanical sources (named in our context as an "occurrence") and made available for statistical purposes.

Several applications of SISTEMAT have appeared in the literature. One of the initial programs that searches characteristic signals in ^{13}C NMR spectra was published in 1990 [23] and was first applied to a sesquiterpene-containing database. The ^{1}H NMR data were exploited in the study of iridoids to find the most likely carbon skeletons [24]. The union of several programs that aid the processes of structural elucidation inside a single module, named MONOREG, was first applied to monoterpenes [25]. This module includes the identification of common groups attached to carbon skeletons [26], the simple similarity search between ^{13}C NMR spectra, the search for some functional groups by ^{13}C NMR [27] and finally the most important program of the system, named SISCONST [28].

After the initial publication, SISCONST was applied to solve problems in various subfields of natural products including sesquiterpenes [29], diterpenes [30], triterpenes [31] and alkaloids [32]. It was also applied to a database of 700 flavonoids in order to check its search performance for non-aliphatic substructures. Despite the great number of signals in the 90-160 ppm range, the algorithm used for terpenoids compounds was able to correctly identify the two monomeric units within a biflavonoid [33].

SISCONST was improved since the times of its early development .The maximum number of input signals in SISCONST was initially limited to 50. A recent version was designed to work with the ^{13}C NMR DEPT of terpenoid mixtures from essential oils. Presently, there is no limit for the number of input signals. The user can also precisely control the tolerance that defines if a group of sample signals matches a group of signals from the database. The corresponding version of SISCONST was used to analyze the essential oil of two species of Piperaceae [34]. The most recent version was used by another research group to identify a clerodane diterpene from *Plectranthus ornatus* [35]. Full technical details were also reported [36]. The program includes a better user interface, developed in Turbo Pascal that provides new facilities for data input, data management and substructure viewing. The core FORTRAN code was freed from its early limitations and rewritten in C language so that is can be now run on various platforms. Programs and databases are available upon request to the author at São Paulo University.

SISCONST performs an analysis of the DEPT spectra of an unknown and produces two lists: the first contains all subspectra-substructure matchings, selected according to user-provided criteria such as the tolerance that defines equality between chemical shifts, the minimum number of atoms for a substructure to be retrieved and the tolerance increment value to apply in order to redo a search when the last one has failed. The second list reports a statistical analysis of all substances within a specific chemical class that have greater similarity with the sample using the largest possible tolerance (5 ppm). The constraints given by SISTEMAT are used within the LSD system as described below.

We will describe hereafter the computer-assisted elucidation of several compounds that were reported in the recent literature, using a combination of LSD and SISTEMAT CASE tools. The examples were selected among terpenic compounds, in increasing order of molecular complexity. The results can be reproduced by the reader as the LSD input files are available from www.univ-reims.fr/LSD/Npc2005.zip. The file names are cited in the text and appear in ***bold italics***.

Example 1 is a rearranged sesquiterpene (Figure 1), photochemically obtained from a natural compound containing the longipinene skeleton.

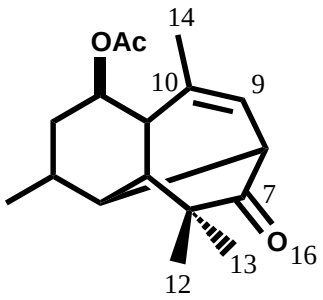

Figure 1: A rearranged longipinene sesquiterpene [37]

The HMBC and HSQC spectra data, the fact that methyl groups 12, 13, 14 are singlets in the ^{1}H NMR spectrum and thus must be bound to quaternary carbons, and the two pairs of sp^2 carbon data were given to LSD, resulting in the generation of 132 solutions (***Rodrigues7a***)**.** The acetyl group, trivially identified in the spectrum, was removed from the data set and replaced by a hydrogen atom. SISTEMAT proposes some substructures and their assignment in the ^{13}C NMR spectrum (Figure 2):

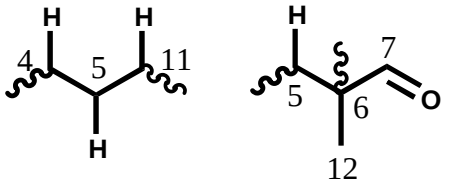

Figure 2: Fragments listed by SISTEMAT after analysis of the DEPT spectra of the sesquiterpene in Figure 1.The atom numbers refer to those given in the literature. The same numbers were used to write the LSD BOND commands.

The two substructures were introduced in LSD as BOND commands (***Rodrigues7b***) and the program gave only one solution as the deacetylated derivative of the compound drawn in Figure 1.

Example 2 is another rearranged sesquiterpene, also photochemically obtained form a natural longipinene. The 2D information for the compound in Figure 3 was analyzed first by LSD to give 26 solutions (***Rodrigues8a***). Once the two fragments C2-C3-C4-C8 and C1-C11-C5, given by SISTEMAT, were taken into account by LSD, a single solution was obtained (***Rodrigues8b***).

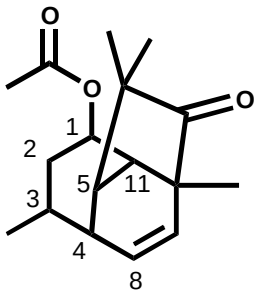

Figure 3: Another rearranged sesquiterpene [37].

Example 3 represents a highly oxidized sesquiterpene lactone (Figure 4). Using only HMQC and HMBC correlations led to millions of solutions (***Kinoshita1a***). The 2D NMR data and the constraints given by the authors of [38], require that four methine carbons, C2-C3-C4-C5, are linked in row and that C2, C3 and C6 are linked to oxygen atoms (***Kinoshita1b***)**.** This results in 57943 solutions. SISTEMAT found a substructure made of two vicinal epoxide groups, thus allowing to the entry of C11-C12-C13-C14 fragment in the LSD input file (***Kinoshita1c***) that yields 400 solutions. Finally, introducing the C7-C1-C6 fragment from SISTEMAT and the oxygens of the epoxide groups led to a single solution (***Kinoshita1d***)**.**

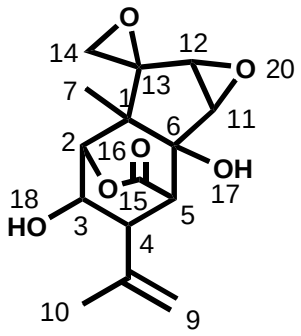

Figure 4: A sesquiterpene lactone isolated from *Coriaria japonica* [38].

Example 4 is neoclerodane diterpene isolated from *Salvia Divinorum*. This is a difficult test for both systems, firstly due the poor description of the published 2D NMR data, and secondly due to the absence of previously reported substructures such as the oxidized side chain of this diterpene [39].

LSD alone gave 180328 solutions (***Harding4a***). When the DEPT spectrum was introduced in SISTEMAT, it indicated that there is a high probability that we are working with a neoclerodane skeleton (***Harding4b***). This information was coded as a substructure, leading to 711 solutions. The ring system in Figure 5 (left), brought from SISTEMAT to LSD, produced 58 solutions (***Harding4c***). Finally, the fragment in Figure 5 (right), also derived from SISTEMAT, lead to a set of 3 structures that contains the published one, drawn in Figure 6 (***Harding4d***).

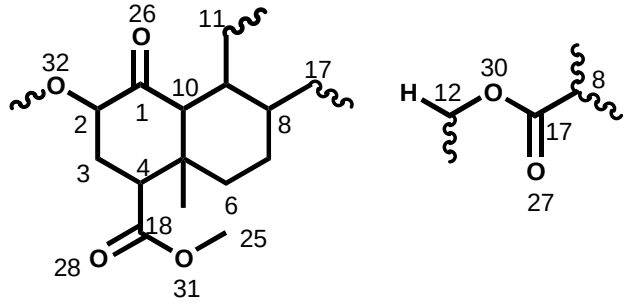

Figure 5: Substructures found by SISCONST for the diterpene of Example 4.

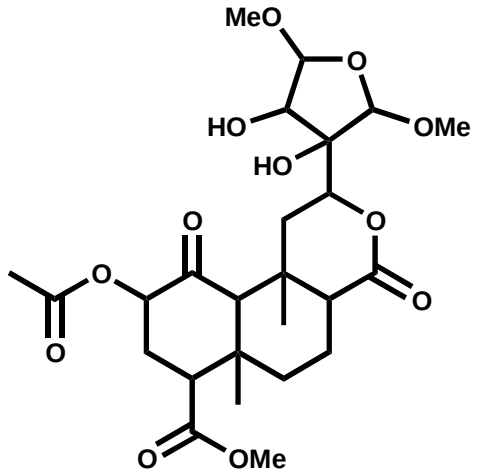

Figure 6: A diterpene isolated from *Salvia divinorum* [39].

The two other structures contain either a hydroperoxide or a carbonate function and can be quickly discarded.

In order to test the two computer systems with larger problems, the spectra of a new steroid ($C_{29}H_{44}O_3$) was selected as Example 5. This compound was recently isolated from the soft coral *Sinularia dissecta* [40]. A first run with LSD (***Jin6a***) gave rise to an extremely severe combinatorial explosion when working with HMBC and HMQC data, and restricting singlet methyls, doublet methyls and deshielded carbons to be bound to quaternary carbons, CH groups, and to oxygen atoms, respectively.

The botanical data base in SISTEMAT indicated that a probability of 46% for the presence of an ergostane steroid skeleton and of 19% for the presence of a cholestane skeleton (that differs only by the side chain that is connected to ring D). The DEPT spectrum, once analyzed by SISCONST, gave probabilities of 39% for cholestane and 19% for ergostane. We tried first to impose on LSD an ergostane skeleton as substructure, and found 54 solutions (***Jin6b***). The assignment of the atoms in the A – B ring system (Figure 7, left) lead to 15 solutions (***Jin6c***). When C-18 was constrained to be carbon of a lactone group (***Jin6d***), two solutions were found. Indeed they both refer to the same molecular structure (Figure 7, right) but differ in their ^{13}C chemical shift assignment.

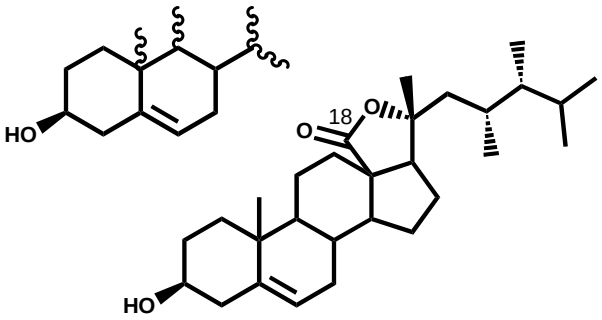

Figure 7: (Left) The A-B ring system and its attached hydroxyl group as given by SISTEMAT. (Right) The solution of Example 5.

Example 6, a new seco-iridoid isolated from *Calycophyllum spruceanum* (Rubiaceae) [41], also showed that constraints from SISTEMAT can facilitate the elucidation process by LSD when the set of correlation data is poor. The data from the sugar part of the molecule was removed, as it can be easily recognized. Working with HMBC data alone, no solution was produced within a reasonable amount of time (***Zuleta3a***). A simple analysis of the DEPT spectrum indicated that at least six bonds can be introduced in LSD as three C=O groups, two carbon-carbon double bounds and a carbon linked to two oxygen atoms. This information leads to 2164 solutions (***Zuleta3b***). SISTEMAT found the six-membered ring of the iridoid skeleton within a sequence of eight bonds, as described in Figure 8 (left).

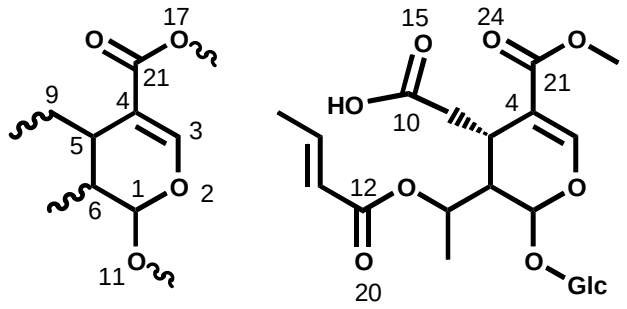

Figure 8: The constraint and the correct solution of the iridoid problem.

With these bonds, two solutions were found (***Zuleta3c***). The correct one is drawn in Figure 8 (right). The other one contains a hydroperoxide group and cannot be considered as likely.

Finally, a triterpene from *Petasites tricholobus* [42] (Figure 9) was chosen as seventh test for the LSD-SISTEMAT association. The compound has a rearranged ursane skeleton that was only recently described and that therefore is obviously not present in the database. LSD, without constraints, gave 24504 solutions (***Xie1a***). With the triterpenic five six-membered ring system as constraint, LSD gave 56

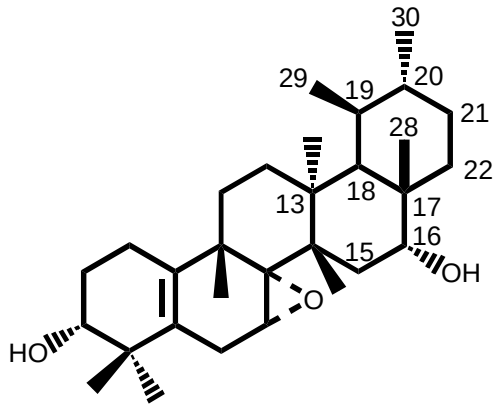

Figure 9: A triterpene isolated from *Petasites tricholobus* [42].

solutions (***Xie1b***). Restricting singlet methyls, doublet methyls and deshielded carbons to be bound to quaternary carbons, CH groups, and to oxygen atoms, respectively, yielded 24 solutions (***Xie1c***). The search in the triterpene database of SISTEMAT led to the assignment of the carbon atoms whose numbers are reported in Figure 9, thus forcing LSD to produce only 6 solutions, among which the published one can be found (***Xie1d***). The latter is the only one that contains an epoxide group and that matches best with the expected ^{13}C chemical shifts for this group.

Structure generation, when carried out on a computer equipped with a Pentium 4 processor at a 2.8 GHz clock frequency, is a matter of a few seconds or much less, unless a combinatorial explosion occurs (as in Kinoshita1a, Zuleta3a, Jin6a and Jin6b LSD input files) .

When LSD provides more than one solution, the problem of the selection of the most likely one is often difficult. LSD sometimes produces structures that can be discarded on simple chemical grounds. In less obvious cases, spectral simulation, from data bases, from artificial neural networks or from quantum mechanical calculations is a useful tool for best solution selection [43].

LSD is a "flat" structure generator and cannot use 3D information provided by either nOe or *J* coupling constant measurements. The general problem of the automatic stereochemical analysis of small molecules is a difficult one, especially when they are conformationally flexible. However, the structure fragments that are provided by SISTEMAT may contain absolution configuration data, because it was exclusively built from the natural product chemistry literature and because stereochemistry is of the highest importance in this field.

The LSD program, when coupled with the natural product oriented data base system SISTEMAT, is able to propose to the user a restricted set of possible planar molecular formulas as solutions of a problem in the field of natural product structure elucidation, even when the quality of the 2D NMR data is poor. The present availability of the LSD software and the future one of SISTEMAT will certainly provide the community of natural product chemists with a valuable tool for increasingly easier structure solving tasks.

Acknowledgments: This work was financed by CNRS (France), Fundação de Amparo a Pesquisa do Estado de São Paulo e CNPq (Brazil).

References

[1] Carhart RE, Smith DH, Gray NAB, Nourse JG, Djerassi C. (**1981**) GENOA: A Computer Program for Structure Elucidation Utilizing Overlapping and alternative Substructures. *Journal of Organic Chemistry*, ***46***, 1708-1718.

[2] Dubois JE, Carabedian M, Dagane I. (**1984**) Computer-aided elucidation of structures by carbon-13 nuclear magnetic resonance. *Analytica Chimica Acta*, ***158***, 217-233.

[3] Carabedian M, Dagane I, Dubois JE. **(1988)** Elucidation by Progressive Intersection of Substructures from Carbon-13 Nuclear Magnetic Resonance. *Analytical Chemistry*, ***60***, 2186-2192.

[4] Shelley CA, Munk ME. (**1981**) . CASE, a computer model of the structure elucidation process. *Analytica Chimica Acta*, ***133***, 507-516.

[5] Christie BD, Munk ME. (**1991**) The Role of Two-dimensional Nuclear Magnetic Resonance Spectroscopy in Computer-Enhanced Structure Elucidation. *Journal of the American Chemical Society*, ***113***, 3750-3757.

[6] Bremser W, Grzonka M**. (1991)** SpecInfo-A Multidimensional Spectroscopic Interpretation System. *Mikrochimica Acta*, ***2***, 483-491.

[7] Will M, Fachinger W, Richert JR. (**1996**) Fully Automated Structure Elucidation – A Spectrocopist's Dream Comes True. *Journal of Chemical Information and Computer Science*, ***36***, 221-227.

[8] Funatsu K, Miyabayashi N, Sasaki S. (**1988**) Further development of structure generation in the Automated Structure Elucidation System CHEMICS. *Journal of Chemical Information and Computer Science*, ***28***, 18-28.

[9] Funatsu K, Susuta Y, Sasaki S**.** (**1989**) Introduction of Two-Dimensional NMR Spectral Infornation to an Automated Structure Elucidation System, CHEMICS. *Journal of Chemical Information and Computer Science*, ***29***, 6-11.

[10] Peng C, Yuan S, Zheng C, Hui Y. **(1994)** Efficient Application of 2D NMR Correlation Information in Computer-Assisted Structure Elucidation of Complex Natural Products. *Journal of Chemical Information and Computer Science*, ***34***, 805-813

[11] Lindel T, Junker J, Köck M. (**1997**) Cocon: From NMR Correlation Data to Molecular Constitutions. *Journal of Molecular Modeling*, ***3***, 364-368.

[12] Elyashberg ME, Blinov KA, Williams AJ, Martirosian ER, Molodtsov SG. (**2002**) Application of a New Expert System for the Structure Elucidation of Natural Products from Their 1D and 2D NMR Data. *Journal of Natural Products*, ***65***, 693-703.

[13] Steinbeck C. (**1996**) A program for Structure Elucidation from NMR Correlation Experiments. *Angewandte Chemie International Edition in English*, ***35***, 1984-1986.

[14] Jaspers M. (**1999**) Computer assisted structure elucidation of natural products using two-dimensional NMR spectroscopy. *Natural Product Reports* ***16***, 241-248.

[15] Steinbeck C. (**2004**) Recent developments in automated structure elucidation of natural products. *Natural Product Reports*, ***21***, 512-518.

[16] Nuzillard J-M, Naanaa W, Pimont S, **(1995)** Applying the Constraint Satisfaction Problem Paradigm to Structure Generation. *Journal of Chemical Information and Computer Science*, ***35***, 1068-1073.

[17] Nuzillard J-M. (**2003**) Automatic Structure Determination of Organic Molecules: Principle and Implementation of the LSD Program. *Chinese Journal of Chemistry*, ***21***, 1263-1267

[18] Ley SV, Doherty K, Massiot G, Nuzillard J-M, (**1994**) "Connectivist" Approach to Organic Structure Determination. LSD-Program Assisted NMR Analysis of the Insect Antifeedant Azadirachtin. *Tetrahedron*, ***50***, 12267-12280.

[19] Alvarenga SAV, Gastmans JP, Rodrigues GV, Emerenciano VP. (**2001**) A computer-assisted approach for chemotaxonomic studies - Diterpenes in Lamiaceae. *Phytochemistry*, ***56***, 583-595.

[20] Emerenciano VP, Alvarenga SAV, Rodrigues GV, Ferreira MJP.(**2005**) A general survey and some taxonomic implications of diterpenes in the Asteraceae. *Journal of The Linnean Society*, ***147***, 291-308.

[21] Emerenciano VP, Alvarenga SAV, Gastmans JP, Rodrigues GV, Ferreira MJP.(**2003**) SISTAX - An Intelligent tool in recovering information on natural products. *Journal of Brazilian Chemical Society*, ***14***, 369-374.

[22] Gastmans JP, Zurita JC, Emerenciano VP. (**1989**) Prévision des Spectres de ^{13}C par Ordinateur. Le Problème de La Codification. *Analytica Chimica Acta*, ***217***, 85-100.

[23] Gastmans JP, Lopes MN, Furlan M, Emerenciano, VP.(**1990**) L' étude des signaux de ^{13}C par microordinateur : un programme simple et rapide. *Computers & Chemistry*, ***14***, 75-78.

[24] Emerenciano VP, Ferreira MJP, Rodrigues GV. (**2004**). Application of a new program, H1MACH, for predicition of iridoids skeletons. *Spectroscopy Letters*, ***37***, 587-605.

[25] Ferreira MJP, Rodrigues GV, Emerenciano, VP.(**2001**) . Monoreg - An expert system for structural elucidation of monoterpenes. *Canadian Journal of Chemistry*, ***79***, 1915-1925.

[26] Ferreira MJP, Oliveira FC, Alvarenga SAV, Macari PAT, Rodrigues GV, Emerenciano VP. (**2002**) Automatic Identification by ^{13}C NMR of substituents bonded in natural products skeletons. *Computers and Chemistry*, ***26***, 601-632.

[27] Ferreira MJP, Rodrigues GV, Emerenciano VP. (**2001**) Automatic identification of terpenoid skeletons through ^{13}C NMR data disfunctionalization. *Analytica Chimica Acta*, ***429***, 151-170.

[28] Fromanteau DL, Gastmans JP, Vestri SA, Emerenciano VP, Borges JHG. (**1993**) A Constraint Generator In Structural Determination By Microcomputer. *Computer and Chemistry*, ***17***, 369-378.

[29] Oliveira FC, Ferreira MJP, Núñez CV, Rodrigues GV, Emerenciano VP. (**2000**). ^{13}C NMR spectroscopy of eudesmane sesquiterpenes. *Progress in NMR Spectroscopy*, ***37***, 1-45.

[30] Vestri SA., Gastmans JP, Rodrigues GV, Emerenciano, VP. (**1997**) Ditregra - An Auxiliary Program for Structural Determination of Diterpenes. *Spectroscopy*, ***13***, 227-249.

[31] Macari PAT, Gastmans JP, Rodrigues GV, Emerenciano, VP**.** (**1994**) An Expert System for Structure Elucidation of Triterpenes. *Spectroscopy*, ***12***, 139-167.

[32] Rufino AR. (**2005**) Using Computers for the Structural Elucidation of Alkaloids (in Portuguese). São Paulo University, São Paulo, Brazil.

[33] Emerenciano VP, Melo DL, Rodrigues GV, Gastmans JP. (**1997**) Application of artificial intelligence in organic chemistry. Part XIX. Pattern recognition and structural determination of flavonoids using ^{13}C NMR spectra. *Spectroscopy*, ***13***, 181-190.

[34] Ferreira MJP, Costantin MB, Sartorelli P, Rodrigues GV, Limberger R, Henriques AT, Kato MJ, Emerenciano VP. (**2001**) Computer-aided method for identification of components in essential oils by ^{13}C NMR spectroscopy. *Analytica Chimica Acta*, ***447***, 125-134.

[35] Oliveira PM, Ferreira AA, Silveira D, Alves RB, Rodrigues GV, Emerenciano VP, Raslan DS. (**2005**) Diterpenoids from the aerial parts of *Plectranthus ornatus*. *Journal of Natural Products*, ***68***, 588-591.

[36] Ferreira MJP. (**2003**) Spectral analysis, structure generation and prediction of ^{13}C NMR spectra (in Portuguese). Doctoral Thesis. São Paulo University, Sao Paulo, Brazil.

[37] Rodrigues MM, Rojas MCG, Nathan PJ. (**2002**) Quirogane, Prenopsane, and Patzcuarane Skeletons Obtained by Photochemically Induced Molecular Rearrangements of Longipinene Derivatives. *Journal of Natural Products*, ***65***, 1398-1411

[38] Kinoshita T, Itaki N, Hikita M, Aoyagi Y, Hitotsuyanagi Y, Takeya K. (**2005**) The Isolation and Structre Elucidation of a New Sesquiterpene lactone from the Poisonous Plant *Coriaria japonica* (Coriariaceae*). Chemical and Pharmaceutical Bulletin **53***, 1040-1042

[39] Harding WW, Tidgewell K, Schmidt M, Shah K, Dersch CM, Snyder J, Parrish D, Deschamps JR, Rothman RB, Prisinzano TE. (**2005**) Salvinicins A and B, New Neoclerodane Diterpenes from *Salvia divinorum*. *Organic Letters*, ***7***, 3017-3020.

[40] Jin P, Deng Z, Pei Y, Fu H, Li J, Van Ofwegen L, Proksch P, Lin W. (**2005**) Polyhydroxylated steroids from the soft coral *Sinularia dissecta*. *Steroids*, ***70***, 487-493.

[41] Zuleta LMC, Cavalheiro AJ, Silva DHS, Furlan M, Young MCM, Albuquerque S, Castro-Gamboa I, Bolzani VS (2005) *seco*-Iridoids from *Calycophyllum spruceanum* (Rubiaceae). *Phytochemistry*, ***64***, 549-553.

[42] Xie WD, Zhang Q, Li PL, Jia ZJ. (**2005**) Two triterpenoids and other constituents form *Petasites tricholobus*. *Phytochemistry*, ***66***, 2340-2345.

[43] Meiler J, Maier W, Will M, Meusinger R. (**2002**) Using neural networks for ^{13}C NMR shift prediction – comparison with traditional methods. *Journal of Magnetic Resonance*, ***157***, 242-252.