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

Computer-aided Identification of Chemical Constituents Isolated from *Cybistax antisyphilitica*

Maria Paula P. Ramos,[1] Marcelo J. P. Ferreira,[2] Lousã Lopes,[1] and Vicente P. Emerenciano,[2,*]

[1] Instituto de Saúde Coletiva, Universidade Federal de Mato Grosso, Cuiabá, MT, Brazil
[2] Instituto de Química, Universidade de São Paulo, Caixa Postal 26077, 05513-970, São Paulo, SP, Brazil

Received xxx; Preprint published xxx; Accepted xxx ; Published xxx

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

Abstract

Motivation. Natural products are substances generally obtained in mixtures from natural sources. For identification and structure determination it is necessary to isolate each compound present in the mixture through chromatographic methods. This procedure is frequently used due to the great diversity and the structural complexity of the compounds found within the mixtures, and usually demands time. So, trying to reduce the time spent during this process, we developed an expert system able to analyze complex mixtures of compounds.

Method. SISCONST program was developed to aid the process of structural determination of organic compounds by means of analysis of chemical shifts and multiplicities obtained from the ^{13}C NMR DEPT spectra. The program may predict the most probable skeleton type for a compound under analysis and suggests several substructures from the signals assigned in ^{13}C NMR. In this study, we have used the program to identify the individual components present in a mixture from the ^{13}C NMR data.

Results. The program was evaluated with a mixture of triterpenes isolated from *Cybistax antisyphilitica*. The results show that the program was able to propose all components present in the mixture, including a triterpene not reported yet in the literature.

Conclusions. The method described in this study allowed us to achieve the correct identification of the substances present in the mixture. The procedure can therefore be applied as a tool for analysis of compounds whose separation through chromatographic methods is very difficult. One of the advantages of this method is to detect the presence of new compounds in mixtures.

Availability. About the software used in this study, its availability or more details can be supplied by getting a touch with the corresponding author.

Keywords. ^{13}C NMR; mixture analysis; computer-assisted method; triterpene; *Cybistax antisyphilitica*.

Abbreviations and notations

^{13}C NMR, Carbon-13 nuclear magnetic resonance	HMQC, Heteronuclear Multiple-Quantum Coherence
EtOAc, ethyl acetate	HMBC, Heteronuclear Multiple Bond Correlation
IR spectra, infrared spectra	DEPT, Distortionless Enhancement by Polarization Transfer

1 INTRODUCTION

The development of expert systems for structural elucidation of new organic compounds has been the objective of various research groups in the last decades [1-3]. However, such systems do not allow the analysis of complex mixtures of substances without previous purification, due to the large number of signals and overlapping of these.

In natural products' chemistry, during extraction phase, it is very common the obtaining of complex mixtures of compounds. For such cases, we have developed a methodology to analyze ^{13}C

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

NMR data of constituents present in mixtures through our expert system SISTEMAT [4,5]. This methodology had been applied in the identification of a mixture of triterpenes obtained from the roots of *Vernonia cognata* (Asteraceae) [6] and had also been evaluated in the identification of components in the essential oils from leaves of *Piper cernuum* and *Piper regnellii* (Piperaceae) [7]. In both cases, the components in the mixture were successfully recognized by the system SISTEMAT. However, the procedure based on analyses of ^{13}C NMR data had not been tested yet in identification of new chemical constituents present in the mixture.

The aim of this paper is to present the structure of a new pentacyclic triterpene, which was identified through the expert system SISTEMAT in a mixture of compounds isolated from *Cybistax antisyphilitica*.

2 MATERIALS AND METHODS

2.1 The Plant Material

The genus *Cybistax* belongs to the family Bignoniaceae, which comprises 120 genera and 820 species [8,9]. For many decades, plants of this whole family have been used owing to their medicinal properties. These properties include several biological activities, such as, antitumoral, dermatological, antimicrobial, purgative, wound healing, and others, e. g. insect antifeeding activities [10-18].

Cybistax antisyphilitica has been used for a long time in popular medicine. It is found growing in the cerrado and is a tree commonly native to Brazil, where it is popularly known as “ipê-branco”, “cinco-folhas” and “pé-de-anta”. The decoction of roots, wood and young leaves of the plant has been used as depurative, antisyphilitic and diuretic agents [19]. A previous phytochemical study on constituents of *Cybistax antisyphilitica* reported the presence of oleanolic acid and iridoids as 6-O-(p-coumaroyl)-catalpol and macfadienoside. Analgesic and anti-inflammatory properties of this plant have also been reported [20] and was demonstrated that the isolated compounds did not show cytotoxic effect on pig kidney cell line IB-RS-2 growth after 72h test [19].

For this study, leaves of *Cybistax antisyphilitica* were collected on April 08, 1998 along the road sides of via Cuiabá-Marzagão, Cuiabá, MT, Brazil and were identified by Prof. Dr. Germano Guarim Neto – Institute of Biological Sciences, Universidade Federal do Mato Grosso. A voucher specimen (no. 18976) was deposited at the Central Herbarium of the Universidade Federal do Mato Grosso, MT, Brazil.

2.2 Extraction and Isolation

Dried and pulverized plant material, 4.43kg, was exhaustively extracted with hexane in Soxhlet, during 12h, affording 240.03g of crude extract. The extract was partitioned first with hexane and then with EtOAc.

The hexane-soluble fraction was submitted to a chromatography process over Si gel 60 and then eluted with hexane, by gradually increasing the polarity of the medium with acetone. After the TLC tests, ten fractions of 125mL were collected and reagrouped in four fractions. The fraction eluted with hexane-acetone 19:1 showed a residue that was rinsed and recrystallized in ethanol, furnishing 50mg of an amorphous white solid that was identified as nonacosane.

The EtOAc-soluble fraction (2.0g) was submitted to a chromatography process over Si gel 60 and then eluted with hexane, by gradually increasing the polarity of the medium with acetone. Seventy-five fractions of 25mL were collected and reagrouped in ten fractions. The fraction eluted with hexane-acetone 75:25 furnished 20mg of β-sitosterol. The fraction eluted with hexane-acetone 50:50 provided a mixture of two triterpenes that were identified through the expert system SISTEMAT.

2.3 The SISCONST Program

The identification of the triterpenes present in the mixture was based on the analysis of their ^{13}C NMR spectra, followed by comparison with those of triterpenes encoded and stored in our database [21]. The program used to analyze the ^{13}C NMR data of the mixtures was the SISCONST [22].

The SISCONST program was developed to assist in the process of structural determination of natural products through the analysis of chemical shifts and multiplicities obtained from the ^{13}C NMR DEPT spectra. The program can predict the most probable skeleton type of a compound under analysis and suggest several substructures with the signals of ^{13}C NMR assigned.

The program matches the signals of the spectrum obtained with all spectra stored in the database. If a spectrum signal and its respective multiplicity are present in a determined carbon atom, the signals of the interlinked carbons are matched with the spectrum under study. 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.

The information to be supplied to the SISCONST program demands the following data: ^{13}C chemical shifts and their multiplicities obtained experimentally from the DEPT spectra; the error range admitted by the program, being generally 1.0δ; the gradient, that allows the automatic increase of the error range; and the minimum number of carbons for the substructure search, i.e. this number is the half of the total carbon number of the studied chemical class.

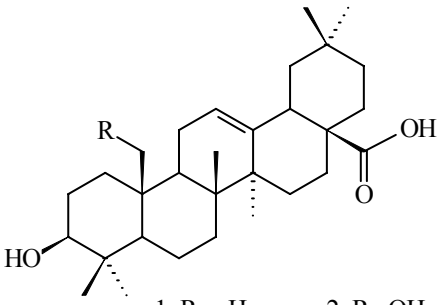
The search process for a probable skeleton is limited to compounds having a substructure

containing at least half of the total carbon number and has been previously detailed [22]. However, for analysis of natural products' mixtures, the skeleton probability determination is not computed due to the diverse skeletal types that can be encountered in a mixture.

3 RESULTS AND DISCUSSION

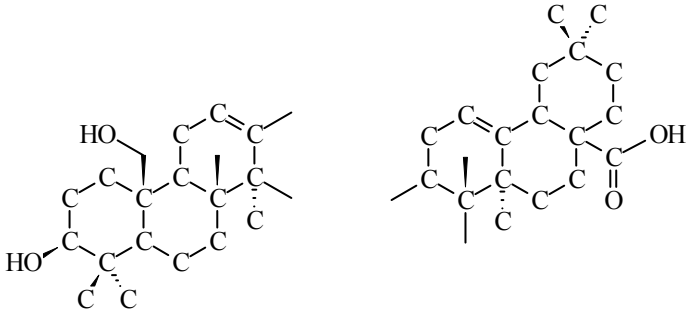
The spectrum of the triterpene mixture was obtained and its signals (43 signals) were used as input to the program SISCONST. The program efficiency was based on the comparison of results obtained in the identification of compounds by analysis of ^{13}C NMR spectra in the mixture with those obtained in the identification of individual compounds by analysis of HMQC and HMBC spectra.

The program matched the signals of the former with those of all spectra stored in the database. At the end of the analysis, the program exhibited the oleanolic acid structure (**1**, Figure 1). The assignment of its ^{13}C NMR signals was in full agreement with the data previously published [23]. Then, in the second step, the program continued the analysis and additionally identified two substructures shown in Figure 2. The overlapping of these substructures furnished the second substance of the triterpene mixture (**2**, Figure 1) that was identified as 25-hydroxy-oleanolic acid, a triterpene not reported yet in the literature.



1- R = H; 2- R= OH

Figure 1. Triterpenes isolated from *C. antisyphilitica*



Substructure I Substructure II

Figure 2. Substructures furnished by the SISCONST program

The reliability of each assignment of the both compounds was obtained by isolation of each component and obtaining of the ^{13}C NMR (DEPT sequence, HMQC and HMBC) spectra of the isolated substances.

The antimicrobial activities of the triterpenes were assessed employing fungi and bacteria. For all fungi used, these compounds and the original extracts showed no antifungal activity. The original extract exhibited inhibitory activity against *Staphylococcus aureus* and *Escherichia coli*, gram-positive and gram-negative bacteria, respectively, both showing inhibition hale of 10mm at 2g/mL of original extract.

4 CONCLUSIONS

The SISCONST program was able to efficiently identify the compounds present in the mixture. The overall goal of this study has been to develop an ability to predict the molecular structure of new compounds present in a mixture jointly with the compounds previously available in literature. In summary, the method here described allows the correct identification of the substances present in mixtures. The method can be suggested as a tool for analysis of compounds whose separation through chromatographic methods is very difficult and time-demanding tasks. One of the advantages of this method is to detect the presence of new compounds in mixtures.

Experimental Section

General Experimental Procedures. IR spectra were recorded on a FTIR-Bomem Model MB100, Michelson series spectrophotometer. ^{1}H and ^{13}C NMR spectra were obtained on Brucker-500 NMR spectrometer. High-resolution mass spectra were obatined on a VG Autospec-Micromass 70eV eletron impact. Metal points were recorded on a Mettler FP 80-Hot Stage.

25-Hydroxy-oleanolic acid: IR (KBr) ν_{max} cm^{-1} 1700; ^{1}H NMR ($CDCl_3$) δ 5.53 (1H, m), 3.41 (1H, m), 3.30 (1H, dd, J=4.0, 14Hz), 1.32, 1.24, 1.04, 1.03, 1.09, 1.00 (each 3H, s, 6x CH_3); ^{13}C NMR ($CDCl_3$) δ 33.8 (C-1), 27.2 (C-2), 79.0 (C-3), 38.8 (C-4), 55.3 (C-5), 18.3 (C-6), 32.4 (C-7), 39.3 (C-8), 47.7 (C-9), 41.0 (C-10), 23.0 (C-11), 122.7 (C-12), 143.6 (C-13), 41.0 (C-14), 27.7 (C-15), 23.4 (C-16), 46.6 (C-17), 41.1 (C-18), 45.9 (C-19), 30.7 (C-20), 33.1 (C-21), 31.9 (C-22), 28.1 (C-23), 15.6 (C-24), 60.4 (C-25), 17.2 (C-26), 25.9 (C-27), 182.7 (C-28), 32.7 (C-29), 23.6(C-30); HRFABMS *m/z* 472.7004 (calcd for $C_{30}H_{48}O_4$, 472.6997).

Oleanolic acid: IR(KBr) ν_{max} cm^{-1} 1705; ^{1}H NMR ($CDCl_3$) δ 5.51 (1H, m), 3.45 (1H, m), 3.32 (1H, dd, J=4.0, 14Hz), 1.30, 1.24, 1.05, 1.04, 1.02, 0.96, 0.91 (each 3H, s, 7x CH_3); ^{13}C NMR ($CDCl_3$) δ 38.9 (C-1), 28.2 (C-2), 78.0 (C-3), 39.4 (C-4), 55.8 (C-5), 18.8 (C-6), 33.3 (C-7), 39.8 (C-8), 48.1 (C-9), 37.4 (C-10), 23.8 (C-11), 122.5 (C-12), 144.8 (C-13), 42.0 (C-14), 28.3 (C-15), 23.8 (C-16), 46.7 (C-17), 42.0 (C-18), 46.7 (C-19), 31.0 (C-20), 34.3 (C-21), 33.3 (C-22), 28.7 (C-23), 16.5 (C-24), 15.5 (C-25), 17.5 (C-26), 26.2 (C-27), 180.2 (C-28), 33.3 (C-29), 23.8(C-30); HRFABMS *m/z* 456.7009 (calcd for $C_{30}H_{48}O_3$, 456.7003).

β-Sitosterol: MP 138.8-141.7°C; IR(KBr) ν_{max} cm^{-1} 3422, 2936, 2850, 1639, 1465, 1390, 1052, 958, 838; ^{1}H NMR ($CDCl_3$) δ 0.65(3H, s, H-18), 0.78(3H, d, J=6.57Hz, H-26), 0.80(3H, d, J=6.54Hz, H-27), 0.91(3H, d, J=6.48Hz, H-21), 0.98(3H, s, H-19), 3.50(1H, m, H-3), 5.35(1H, m, H-6); ^{13}C NMR ($CDCl_3$) δ 37.6 (C-1), 32.8 (C-2), 72.1 (C-3), 42.6 (C-4), 141.0 (C-5), 122.0 (C-6), 32.0 (C-7), 32.2 (C-8), 50.4 (C-9), 36.5 (C-10), 21.4 (C-11), 37.6 (C-12), 42.7 (C-13), 57.1 (C-14), 24.7 (C-15), 28.6 (C-16), 56.3 (C-17), 12.2 (C-18), 19.4 (C-19), 40.1 (C-20), 20.2 (C-21), 34.1 (C-22), 29.5 (C-23), 46.2 (C-24), 32.3 (C-25), 20.2 (C-26), 19.2 (C-27), 24.7 (C-28), 12.2 (C-29); HRFABMS *m/z* 456.7009 (calcd for $C_{30}H_{48}O_3$, 456.7003).

Nonacosane: MP 62.8-63.1°C; IR(KBr) ν_{max} cm^{-1} 2956, 2917, 2848, 1473, 1462, 1378, 729, 719; ^1H NMR ($CDCl_3$) δ 0.89 (CH_3), 1.26 (CH_2); ^{13}C NMR ($CDCl_3$) δ 14.1(CH_3), 22.7(α-CH_2), 31.9 (β-CH_2), 29.4 (γ-CH_2), 29.7 (others CH_2); FABMS *m/z* 409(M+), 85(36.8), 71(59), 57(100), 43(76.8).

Acknowledgment

This work was supported by grants from the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) and by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). The authors thank Antônio J.C. Brant for helpful discussion during the preparation of the manuscript.

5 REFERENCES

[1] R. K. Lindsay, B. G. Buchanan, E. A. Feigenbaum and J. Lederberg, *Applications of Artificial Intelligence for Organic Chemistry: the DENDRAL project*, McGraw-Hill, New York, 1980.

[2] M. Jaspars, Computer assisted structure elucidation of natural products using two-dimensional NMR spectroscopy, *Nat. Prod. Rep*. **1999**, *16*, 241-247.

[3] M. Badertscher, A. Korytko, K. -P. Schulz, M. Madison, M. E. Munk, P. Portmann, M. Junghans, P. Fontana and E. Pretsch, Assemble 2.0: a structure generator, *Chemom. Intel. Lab. Syst.* **2000**, *51*, 73-79.

[4] J. P. Gastmans, M. Furlan, M. N. Lopes, J. H. G. Borges and V. P. Emerenciano, A inteligência artificial aplicada à química de produtos naturais. O programa SISTEMAT. Parte I - Bases teóricas, *Química Nova* **1990**, *13*, 10-16.

[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, Applications d'intelligence artificielle dans la chimie organique. XVII. Nouveaux programmes du projet SISTEMAT, *Spectroscopy – An International Journal* **1994**, *12*, 91-98.

[6] V. P. Emerenciano, N. F. Roque, M. Furlan and L. M. B. Torres, Computer-aided carbon-13 nuclear magnetic resonance spectroscopy for identification of terpenoid mixtures, *Anal. Chim. Acta* **1990**, *236*, 501-504.

[7] M. J. P. Ferreira, M. B. Costantin, P. Sartorelli, G. V. Rodrigues, R. Limberger, A. T. Henriques, M. J. Kato and V. P. Emerenciano, Computer-aided method for identification of components in essential oils by ^{13}C NMR spectroscopy, *Anal. Chim. Acta* **2001**, *447*, 125-134.

[8] P. Correa, *Dicionário de plantas úteis do Brasil*, Ministério da Agricultura, Rio de Janeiro, 1931.

[9] A. Cronquist, *An integrated system of classification of flowering plants*, Columbia University Press, New York, 1981.

[10] K. C. G. De Moura, F. S. Emery, C. Neves-Pinto, M. C. F. R. Pinto, A. P. Dantas, K. Salomão, S. L. de Castro and A. V. Pinto, Trypanocidal activity of isolated naphthoquinones from *Tabebuia* and some heterocyclic derivatives: A review from an interdisciplinary study, *J. Braz. Chem. Soc.* **2001**, *12*, 325-338.

[11] T. M. A. Alves, A. F. Silva, M. Brandão, T. S. M. Grandi, E. F. A. Smânia, A. S. Junior and C. L. Zani, Biological screening of Brazilian medicinal plants, *Mem. Inst. Oswaldo Cruz* **2000**, *95*, 367-373.

[12] M. A. Rasadah, P. J. Houghton, A. Raman and J. R. S. Hoult, Lipoxygenase inhibitors from Bignoniaceae plants from Malaysia, *J. Trop. Forest Prod.* **1999**, *5*, 71- 79.

[13] D. T. Ferreira, P. S. Alvares, P. J. Houghton and R. Braz-Filho, Constituintes químicos das raízes de *Pyrostegia venusta* e considerações sobre a sua importância medicinal, *Quim. Nova* **2000**, *23*, 42-46.

[14] S. J. Jackson, P. J. Houghton, S. Retsas and A. Photiou, In vitro cytotoxicity of norviburtinal and isopinnatal from *Kigelia pinnata* against cancer cell lines, *Planta Med*. **2000**, *66*, 687-784.

[15] F. G. G. Miranda, J. C. Vilar, I. A. N. Alves, S. C. H. Cavalcanti and Â. R. Antoniolli, Antinociceptive and antiedematogenic properties and acute toxicity of *Tabebuia avellanedae Lor. ex Griseb.* inner bark aqueous extract, *BMC Pharmacol*. **2001**, *1*, 6-10. http://www.biomedcentral.com/1471-2210/1/6

[16] S. Ueda, T. Umemura, K. Dohguchi, T. Matsuzaki, H. Tokuda, H. Nishino and A. Iwashima, Production of anti-tumour-promoting furanonaphthoquinones in *Tabebuia avellanedae* cell cultures, *Phytochemistry* **1994**, *36*, 323-325.

[17] B. Onegi, C. Kraft, I. Kohler, M. Freund, K. Jenett-Siems, K. Siems, G. Beyer, M. F. Melzig, U. Bienzle and E. Eich, Herbal remedies traditionally used against malaria part 6 - Antiplasmodial activity of napthoquinones and one anthraquinone from *Stereospermum kunthianum*, *Phytochemistry* **2002**, *60*, 39-44.

[18] O. A. Binutu and B. A. Lajubutu, Antimicrobial potentials of some plant species of the Bignoniaceae family, *Afr. J. Med. Sci*. **1994**, *23*, 269-73.

[19] J. D. Felício, A. P. Lins, I. C. Simoni and C. R. Gonçalves, Constituents of *Cybistax antisiphylitica*, *Fitoterapia* **1994**, *65*, 281-282.
[20] L. F. V. Bresciani, J. P. Priebe, R. A. Yunes, J. D. Magro, F. D. Monache, F. Campos, M. M. Souza and V. Cechinel-Filho, Pharmacological and phytochemical evaluation of *Adiantum cuneatum* growing in Brazil, *Z. Naturforsch*. **2003**, *58c*, 191-194.
[21] P. A. T. Macari, J. P. Gastmans, G. V. Rodrigues and V. P. Emerenciano, An expert system for structure elucidation of triterpenes, *Spectroscopy – An International Journal* **1994**, *12*, 139-167.
[22] D. L. G. Fromanteau, J. P. Gastmans, S. A. Vestri, V. P. Emerenciano and J. H. G. Borges, A constraints generator in structural determination by microcomputer, *Comp. & Chem.* **1993**, *17*, 369-373.
[23] P. K. Agrawal and D. C. Jain, ^{13}C NMR spectroscopy of oleanane triterpenoids, *Progress in NMR Spectroscopy* **1992**, *24*, 1-90.