

Finding the needle in the haystack. The dereplication of natural product extracts*

Geoffrey A. Cordell† and Young Geun Shin

Department of Medicinal Chemistry and Pharmacognosy, College of Pharmacy,
University of Illinois at Chicago, Chicago, IL 60612, USA

Abstract: Dereplication is an integral part of the early stages of the drug discovery process, where there is a focus on the rapid identification of novel, biologically active metabolites from natural sources. We describe a strategy and a system for dereplication and offer several examples of its successful application.

INTRODUCTION

The development of high-throughput screening procedures for the discovery of biologically active compounds has resulted in a need to prioritize those samples which are deemed 'active'. When these samples are derived from natural product sources (plants, marine organisms or microbes), it is estimated that it takes \$50 000 and 3 months of work to isolate and characterize such an active compound [1]. If the isolate is a well-known, active compound, there will be a distinct sense of 'wasted' effort [2].

To overcome these issues one can limit the number of active extracts entering the fractionation process through a secondary (*in vivo* or *in vitro*) biological screen, which may serve to characterize the activity from a mechanistic or other relevant perspective. Alternatively, one may look chemically at the extract or use chromatographic comparison methods to focus on trying to identify classes of compounds of interest or to be avoided. Under optimum conditions such studies might even be able to indicate whether a biological response is associated with a particular class of compound (bioautography). These dereplication processes do not permit critical decision making based on a probability that the active isolate will be new. For example, they may be focussed on biology and provide only a chemical class inference [3], or they may provide chemical information which cannot be linked to biological activity [4].

We describe here a system which is capable of integrating the physical and the chemical data with biological activity data on well-defined subfractions of an active extract. After injection and crude separation of the extract using a standard gradient elution profile, the absorbance characteristics are obtained with a photodiode array detector. The eluent is split unequally, with the bulk going to 96-well plates while the remainder enters the electrospray mass spectrometer after priming for positive or negative ion analysis. This mass information is then correlated with the biological information obtained on the concurrently collected fractions. In this way, spectral and biological information on the active principle(s) are directly linked in a time frame to specific wells. With this information in hand, there is the need to relate the genus and species under investigation to prior chemical and biological work, and to correlate this with information on those metabolites, from whatever source, which have previously displayed activity. The goal therefore is a system which can afford a direct indication of whether a particular biologically active extract has a high or a low probability that the active principle is a new or a known metabolite [5].

*Invited Lecture presented at the 21st IUPAC International Symposium on The Chemistry of Natural Products (ISCNP-21), Beijing, China, 11–16 October 1998, pp. 1025–1166.

†Corresponding author: E-mail: Cordell@uic.edu

EXPERIMENTAL

Dereplication methodology

HPLC-electrospray/MS

The HPLC-electrospray/MS system for the dereplication consisted of a Hewlett Packard (Wilmington, DE, USA) 5989B mass spectrometer, a 59987A electrospray interface, a Hitachi HPLC system (Hitachi, Japan) with a L-7400 UV detector, a LC Packer 1:50 splitter, and a Gilson F-204 fraction collector (Middleton, WI, USA). The quadrupole temperature of the mass spectrometer was set to 120 °C and the spectra were acquired in the negative or positive ion mode. The temperature of the drying N₂ gas was 350 °C and the flow rate of nebulizing N₂ was 40 mL/min and was maintained at 80 psi for the N₂ stream in the ion chamber.

Chromatographic parameters

The HPLC separation was performed with a Kromasil C-18 reverse phase column (250 × 3.2 mm i.d.) with 5 μm packing material (Technikron, Wilmette, IL, USA). The mobile phase system consisted of water (A) and acetonitrile (B) and the gradient was set up as: 0–1 min 95% A, a linear gradient from 1 to 8 min to 50% A/50% B, a linear gradient from 8 to 20 min to 100% B, and then held at this concentration for a final 10 min before returning to starting conditions. The UV detector was set to 280 nm or 254 nm. The eluent was split 1:50 after the UV detector, and the smaller stream was mixed with postcolumn solvent (methanol:water:triethylamine; 90:9.8:0.2) prior to introduction into the mass spectrometer. The larger stream was recovered in 96-well bioassay plates on a Gilson F-204 series fraction collector (Middleton, WI, USA) and then tested for cytotoxic activity against various human cancer cell lines (Fig. 1).

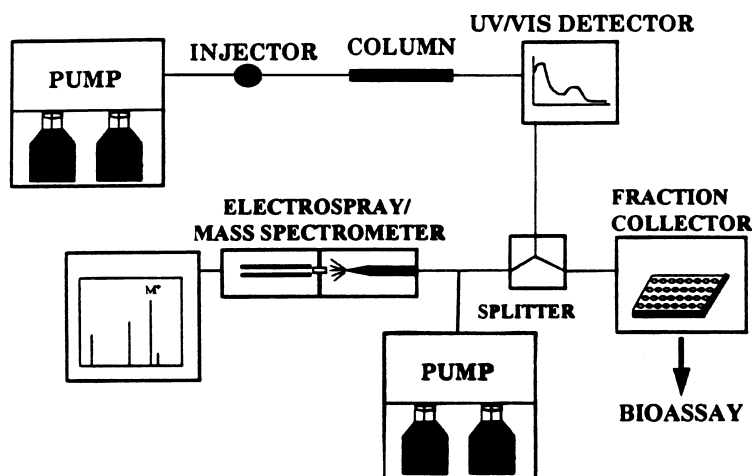


Fig. 1 LC-MS dereplication system.

Evaluation of cytotoxic activity

Extracts or compounds isolated from active plants were tested according to standard procedures [6] against the BC1 (human breast cancer), Lu1 (human lung cancer), Col2 (human colon cancer), LNCaP (human prostate cancer), KB (human oral epidermoid carcinoma), KB-V1 (KB in the presence and absence of vinblastine (1 μg/mL), and ASK (rat glioma cancer) cell systems.

RESULTS AND DISCUSSION

The dereplication process is based on a comparative analysis of datasets already developed with libraries of chemotaxonomic, pharmacological, and molecular weight information, such as NAPRALERT or the *Dictionary of Natural Products* [5]. After the bioassay is completed, masses associated with the biologically active time range(s) are determined from either positive or negative total ion chromatogram (TIC). The time profile of the masses in the area of activity is retrieved from the TIC profiles as extracted ion chromatograms (EIC). The examples illustrate aspects of this dereplication method.

Semecarpus anacardium

Extracts of the fruits of *S. anacardium* L. (Anacardiaceae), collected in India, were evaluated in the human cancer cell panel. A subfraction (SAPE-3) of the petroleum-ether extract of *S. anacardium* was active in the KB and Col2 cell lines with IC_{50} values 2.5 and 5.4 $\mu\text{g/mL}$, respectively, and was applied to dereplication. Figure 2 displays the negative total ion chromatogram (A), the HPLC chromatogram at UV 280 nm (B), the extracted ion chromatograms at m/z 315 and 317 (C), and the cytotoxicity profile (D) using the KB cancer cell system of SAPE-3. There is one area of strong activity and another area with weaker activity in the cytotoxic profile. On the basis of the LC/MS chromatogram, the active components appear at retention times of 15.2 and 19.3 min. The former active compound had a pseudomolecular ion at 316 amu, while the second one displays an ion at 318 amu. These compounds also showed strong absorbance in the HPLC chromatogram at UV 280 nm. The data obtained were correlated from a mass and genus perspective with information in the NAPRALERT database and in the *Dictionary of Natural Products*. Through comparison with these datasets (uv, ms, molecular formula), the active principles were considered to be 1,2-dihydroxy-3-pentadeca-7',10'-dienylbenzene (**1**) and 1,2-dihydroxy-3-pentadec-8'-enylbenzene (**2**), respectively (Fig. 3) [7,8]. These two compounds were also isolated directly from the active extract using semiprep HPLC since the chromatographic properties of the active principles were well-defined. Spectroscopic analysis indicated them to possess the structures indicated above. The two pure isolates were evaluated against the human cancer cell lines and showed significant cytotoxic activities (Table 1).

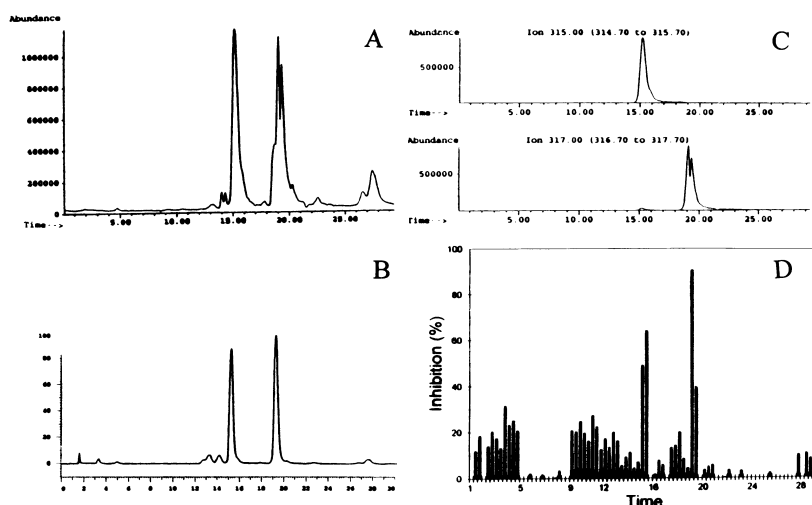


Fig. 2 The negative TIC (A), HPLC chromatogram (B), the EIC (C) and the activity profile (D) against KB of *S. anacardium*.

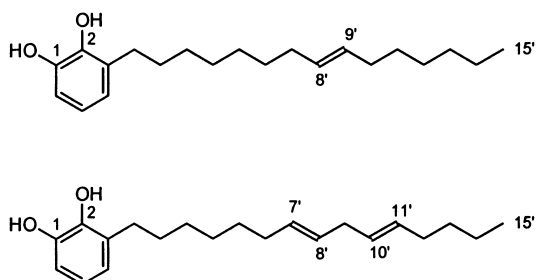


Fig. 3 Cytotoxic compounds in *S. anacardium*.

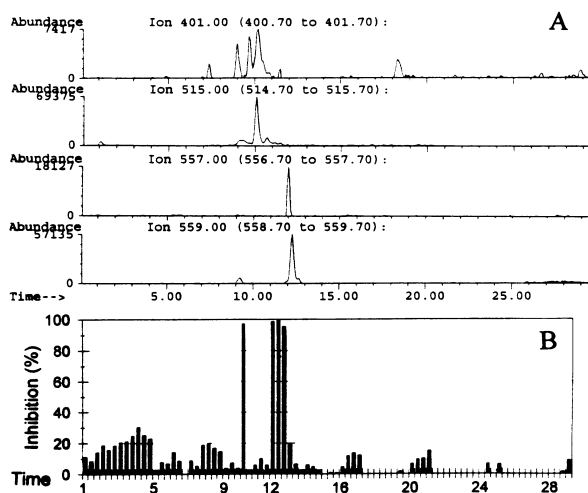
Begonia parviflora

The roots of *B. parviflora* L. (Begoniaceae) were collected in Ecuador and also showed strong cytotoxicity against various human cancer cell lines. In the activity profile of the chloroform extract of

Table 1 Cytotoxic activity of extracts and compounds isolated from *Semecarpus anacardium* and *Acnistus arborescens*

Compound	Cell line tested (ED_{50} , $\mu\text{g/mL}$)						
	BC1	Lu1	Col2	KB	KB-VI ⁺	KB-VI ⁻	LNcaP
SAPE-3 of <i>S. anacardium</i>	>20	13.8	5.4	2.5	10.7	5.3	1.1
1,2-Dihydroxy-3-pentadeca-7',10'-dienylbenzene	8.6	1.9	2.6	1.0	16.7	12.9	>20
1,2-Dihydroxy-3-pentadec-8'-enylbenzene	5.5	2.1	>20	2.7	15.0	13.4	>20
<i>A. arborescens</i>	—	—	—	0.25	—	—	—
Withanolide 1	0.2	1.3	0.03	0.4	0.4	0.4	0.19
Withanolide 2	0.4	2.1	0.08	0.5	0.3	0.4	0.19

B. parviflora, two distinct areas of biological activity were observed, and the retention times of each area were 9.8 min and 12.5 min, respectively. The ions found in the negative extracted ion chromatogram were at m/z 401, 515, 557 and 559 (Fig. 4), which means that the molecular weights of the active compounds would correspond to 402, 516, 558 and 560. These molecular masses were also used to compare with the molecular weight information from the NAPRALERT database.

**Fig. 4** The extracted ion chromatogram (A) and the activity profile (B) against KB of *B. parviflora*.

All of these molecular weights were matched for this *Begonia* sp. in the NAPRALERT system, namely, cucurbitacin D (m.w. = 516), hexanorcucurbitacin D (m.w. = 402), cucurbitacin B (m.w. = 558) and dihydrocucurbitacin B (m.w. = 560), respectively (Fig. 5), and these compounds were previously reported to possess cytotoxic activity [9,10]. After co-injection of authentic samples into the HPLC/MS dereplication system, these compounds were identified as the active principles of *B. parviflora*.

Acnistus arborescens

Acnistus arborescens L. (Solanaceae) was collected in Brazil and also evaluated against the human cancer cell lines. The ether extract was active in the KB system with an IC_{50} value of 0.25 $\mu\text{g/mL}$. Figure 6 is the negative TIC (A) and HPLC chromatogram at UV 254 nm (B), the extracted ion chromatogram (C) and the activity profile (D) against KB of the ether extract of *A. arborescens*. From the activity profile, retention times of areas of strong activity were found to be at 9.3 min, 11.0 min, and 12.4 min, respectively.

The ions found in the negative extracted ion chromatogram (EIC) of the active regions were at m/z 469, 527, 569 and 585, which means that their molecular weights would correspond to 470, 528, 570 and 586. Consequently, these molecular masses were correlated with the information obtained from the NAPRALERT database.

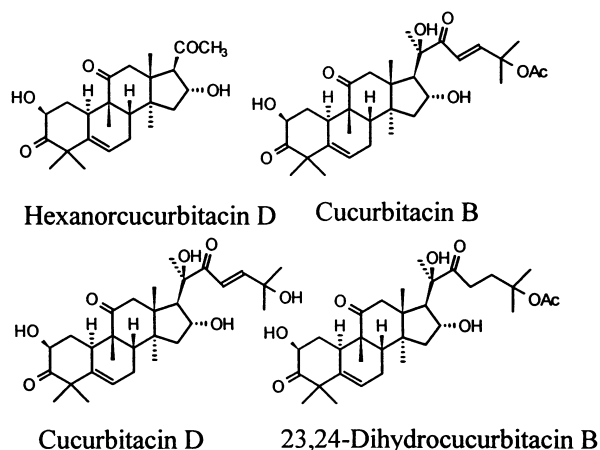


Fig. 5 The cytotoxic compounds identified in *B. parviflora*.

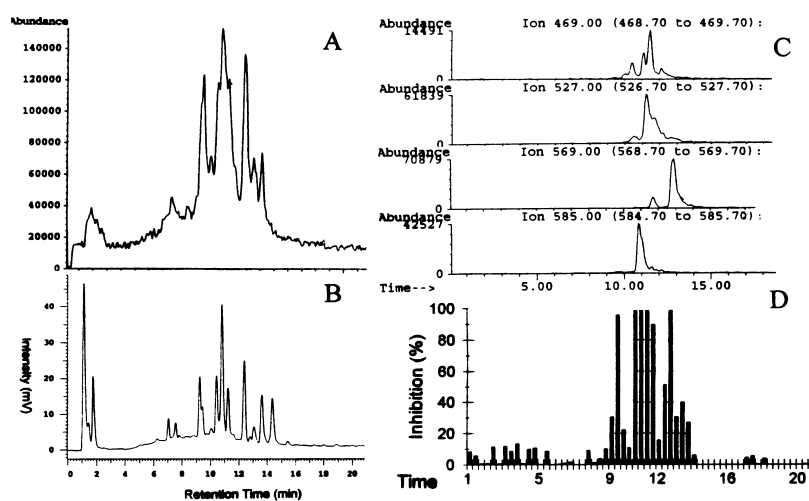


Fig. 6 The negative TIC (A), HPLC chromatogram (B), the EIC (C) and the activity profile (D) against KB of *A. arborescens*.

The only molecular weight that matched with compounds in NAPRALERT was 470, which was identified to be withaferin A; the other masses were not matched in this database. After isolation and structure elucidation, withanolide 1 (m.w. = 528) and withanolide 2 (m.w. = 586) (Fig. 7) were obtained and found to be in the same regions in the activity profile. Both compounds showed strong cytotoxic activity against various cancer cell lines (Table 1). Stereochemical aspects of these compounds are being finalized.

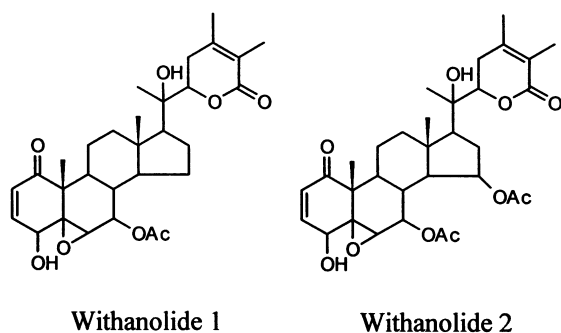


Fig. 7 Cytotoxic compounds from *A. arborescens*.

ACKNOWLEDGEMENTS

This work was supported, in part, by a National Cooperative Natural Product Drug Discovery Grant (CA 52956, Dr A. D. Kinghorn, Principal Investigator). We also thank our collaborators for the provision of extracts, *S. anacardium* (Dr A. V. N. Appa Rao, Kakatiya University, Warangal, India) and *A. arborescens* (Prof. Lauro Barata, Institute of Chemistry-UNICAMP, Campinas, Brazil). Dr J. M. Pezzuto and his associates at UIC are thanked for the bioassay data.

REFERENCES

- 1 D. G. Corley, R. C. Durley. *J. Nat. Prod.* **57**, 1484 (1994).
- 2 G. A. Cordell. *Phytochemistry* **40**, 1585 (1995).
- 3 S. Grabley, J. Wink, A. Zeeck. In *Biotechnology Focus 3: Fundamentals, Applications, Information* (R. K. Finn, P. Prave, M. Schlingmann, W. Crueger, K. Esser, R. Thauer, F. Wagner, eds), pp. 360. Hanser Publishers, Munich (1992).
- 4 G. B. Marini-Bettolo, M. Nicoletti, M. Patamia, C. Galeffi, I. Messana. *J. Chromatogr.* **213**, 113 (1981).
- 5 G. A. Cordell, C. W. W. Beecher, A. D. Kinghorn, J. M. Pezzuto, H. L. Constant, L. Fang, E.-K. Seo, L. Long, B.-L. Cui, K. S. Barrilos. In *Studies in Natural Products Chemistry* (Atta-ur-Rahman, ed.), pp. 749. Elsevier Science Publishers, Amsterdam, the Netherlands (1997).
- 6 K. Likhitwitayawuid, C. K. Angerhofer, G. A. Cordell, J. M. Pezzuto. *J. Nat. Prod.* **56**, 30 (1993).
- 7 P. H. Gedam, P. S. Sampathkumaran, M. A. Sivasambam. *Phytochemistry* **13**, 513 (1974).
- 8 J. A. Hembree, C.-J. Chang, J. L. McLaughlin, G. Peck, J. M. Cassady. *Lloydia* **41**, 491 (1978).
- 9 R. W. Doskotch, M. Y. Malik, J. L. Beal. *Lloydia* **32**, 115 (1969).
- 10 R. W. Doskotch, C. D. Hufford. *Can. J. Chem.* **48**, 1787 (1970).