Journal of Industrial Microbiology & Biotechnology (2000) 25, 342–345

www.nature.com/jim

LC-NMR: a new tool to expedite the dereplication and identification of natural products

SC Bobzin[1], S Yang[2] and TP Kasten[3]

[1]*Monsanto Nutrition and Consumer Products, Monsanto Company, St. Louis, MO 63141;* [2]*Monsanto Analytical Sciences Center, Monsanto Company, St. Louis, MO 63167;* [3]*Searle Discovery Research, Monsanto Company, St. Louis, MO 63167, USA*

The rapid identification of known or undesirable compounds from natural products extracts — ''dereplication'' — is an important step in an efficiently run natural products discovery program. Dereplication strategies use analytical techniques and database searching to determine the identity of an active compound at the earliest possible stage in the discovery process. In the past few years, advances in technology have allowed the development of tandem analytical techniques such as liquid chromatography mass spectrometry (LC-MS), LC-MS-MS, liquid chromatography nuclear magnetic resonance (LC-NMR), and LC-NMR-MS. LC-NMR, despite its lower sensitivity as compared to LC-MS, provides a powerful tool for rapid identification of known compounds and identification of structure classes of novel compounds. LC-NMR is especially useful in instances where the data from LC-MS are incomplete or do not allow confident identification of the active component of a sample. LC-NMR has been used to identify the marine alkaloid aaptamine as the active component in an extract of the sponge *Aaptos* sp. This extract had been identified as an enzyme inhibitor by a high throughput screening (HTS) effort. Isolated aaptamine exhibited an IC_{50}=120 μM against this enzyme. Strategies for the identification of aaptamine and for the use of LC-NMR in a natural products HTS program are discussed. *Journal of Industrial Microbiology & Biotechnology* (2000) **25,** 342–345.

Keywords: natural products; dereplication; LC-NMR; aaptamine

Introduction

The ability to identify known or undesirable compounds rapidly is a critical step in an efficiently run natural products discovery program. This process, commonly called dereplication, is important to prevent unnecessary use of resources on the isolation of known or undesirable compounds from extracts identified by the screening process. This prevents wasted efforts on samples with no potential for development and allows resources to be focused on the most promising leads. The recent application of high throughput screening (HTS) technologies to assay natural products extracts for biological activity has intensified the need for appropriate dereplication strategies. The number of samples identified and the timelines dictated by modern HTS programs require a rapid and effective prioritization of extracts.

Dereplication strategies employ a combination of separation sciences, spectroscopic methods, and database searching [1]. High-performance liquid chromatography (HPLC) has been, by far, the most useful tool for the separation of complex mixtures of small molecules. Reversed phase HPLC on octadecylsilane (ODS or C_{18}) has come to be recognized as the most broadly applicable of bonded phases for this purpose. When interfaced with a diode array detector (DAD), HPLC allows an analyst to identify known compounds by comparison of their HPLC retention time and UV spectra. Unfortunately, this required the compilation of an internal database because these parameters are not present as searchable fields in most commercially available databases. More recently, the advent of electrospray mass spectrometry (ES-MS) provided a MS interface which was compatible with liquid chromatography (LC). In the past 5 years, LC-MS has become a widely used tool for dereplication of natural products [5,7]. LC-MS has become the dereplication tool of choice because the nominal molecular weight can be used as a search query in nearly all databases. Unfortunately, database searching using only the molecular weight produces large answer sets and rarely results in a definitive identification. LC-MS is often combined in series with a DAD, thereby providing UV data to narrow down the answer set. These data, along with the taxonomic information of the source organism, if available, then provide adequate information to narrow the answer set to a short list, or sometimes a single answer. But in many cases, more information is required for a confident identification.

Recently, advances in nuclear magnetic resonance (NMR) spectroscopy have allowed HPLC to be practically interfaced directly with NMR. These advances include the use of higher field magnets (500 MHz and greater) and digital signal processing which have helped address the lack of sensitivity of this technique. In addition, new probe designs, which allow the use of gradient pulse sequences, now provide efficient and specific suppression of the NMR signals due to the HPLC solvents. NMR spectral data provide a great deal of structural information about a compound of interest. The NMR signal for each proton in a molecule provides structural information about the environment and the coupling partners of that proton. Therefore, NMR easily can be capable of discerning structural differences between compounds of the same molecular weight (isobars) or even the same molecular formula (isomers).

Correspondence: Dr SC Bobzin, Searle Discovery Research, Monsanto Company, Monsanto Nutrition and Consumer Products, 800 N. Lindbergh Blvd., Mail Zone T4L, St. Louis, MO 63141, USA. E-mail: Scbobzin@yahoo.com
Received 30 March 2000; accepted 3 July 2000

Materials and methods

Collection and extraction

The marine sponge *Aaptos* sp. was collected by SCUBA at a depth of 13 m offshore of Manado, Northern Sulawesi, Indonesia. The sponge tissue was frozen for storage and then lyophilized. A 1-g portion of the lyophilized tissue was powdered in a Kleco pulverizer and the powder extracted in dichloromethane (50 ml) overnight. The suspension was filtered and the solvent removed to yield an oil which was resuspended in DMSO at 30 mg/ml for screening. This solution was used for all subsequent experiments.

HTS enzyme assay

The HTS assay was formatted in 96-well microtiter plates using a radiometric readout of enzyme activity. Six-point dilution curves of active samples were made starting with 1 μl of extract into a final assay volume of 50 μl. Appropriate controls were performed in each experiment.

LC-MS

LC-MS experiments were performed using a Hewlett Packard 1100 HPLC interfaced with a Perkin Elmer Sciex AP100 mass spectrometer. The chromatography was performed on an Alltech Alltima C_{18} (5 μm) HPLC column (4.6×150 mm) using a solvent gradient from 10% to 100% acetonitrile (0.01% trifluoroacetic acid) in water (0.01% trifluoroacetic acid) in 25 min at a flowrate of 1 ml/min. A 20-μl injection of the crude extract was made and the flow was split (9:1) after the UV detector, with 10% going to the mass spectrometer and 90% collected in a deep dish microtiter plate (1 fraction/min). The fractions were dried on a Savant SpeedVac and then assayed for enzyme activity. Fraction no. 12 possessed the majority of the activity in the HPLC bioassay profile. This activity corresponded to a peak at 10.3 min in the MS chromatogram due to the delay between detectors and the fraction collector.

LC-NMR

The LC-NMR data were acquired using a Varian Unity Inova 500 MHz spectrometer equipped with $^1H\{^{13}C\}$ pulsed field gradient (PFG) LC-NMR flow probe with a 60-μl flow cell (active volume). ^1H NMR spectra were obtained in stop flow mode as described in the *Results* section. Varian WET solvent suppression [8] and related sequences were used to suppress the acetonitrile, its ^{13}C satellites, and the residual water peaks. The WET technique used a series of variable tip-angle solvent-selective RF pulses, where each selective RF pulse is followed by a dephasing field gradient pulse. Free induction decays (FIDs) were collected with 16 K data points, a spectral width of 8000 Hz, a 3-μs 90° pulse, a 2-s acquisition time, and a 1-s pulse delay. A total of 256 transients (about 15 min acquisition time) were acquired to obtain the data in Figure 3. Prior to Fourier transformation, an exponential apodization function was applied to the FID corresponding to a line broadening of 0.25 Hz. The HPLC system consisted of a Varian 9012 solvent delivery system and a Varian 9050 variable wavelength UV–Vis detector. The outlet of the UV detector was connected via a sampling unit (Valco valve) to the LC-NMR flow probe. The HPLC method used an Alltech Alltima C_{18} (5 μm) HPLC column (4.6×150 mm) using a solvent gradient from 10% to 100% acetonitrile (0.01% trifluoroacetic acid) in D_2O (0.01% trifluoroacetic acid) in 25 min at a flowrate of 1 ml/min. A 100-μl

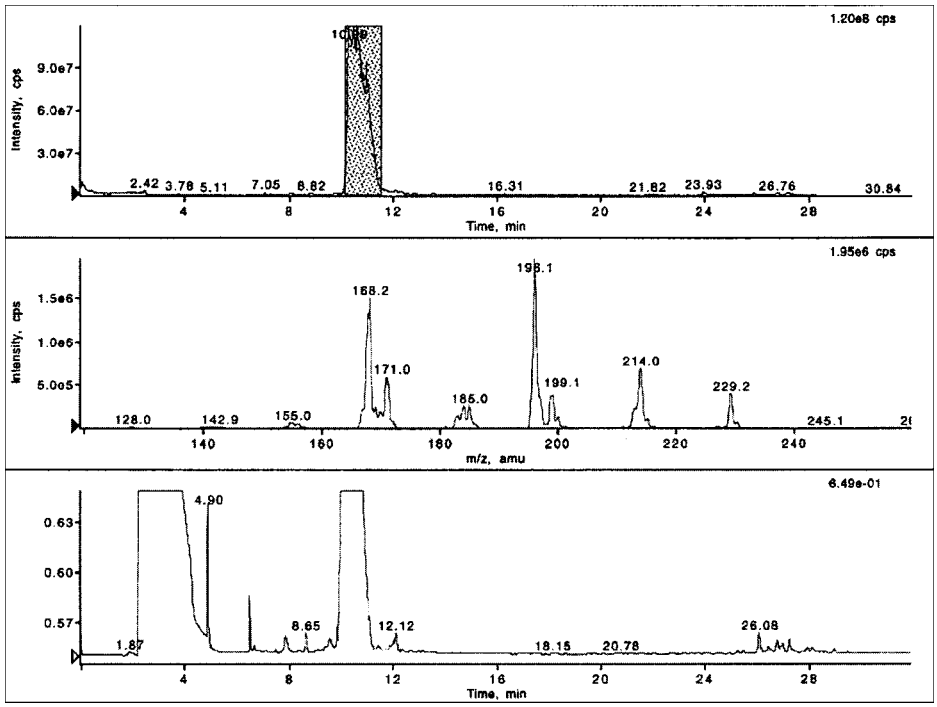

Figure 1 LC-MS data from the separation of the crude extract of *Aaptos* sp. by reversed phase C_{18} HPLC (see text for details). Top panel: Positive ion total ion chromatogram (TIC). Middle panel: Averaged, background subtracted, and smoothed mass spectrum of the peak at 10.3 min [$(M + H)^+$ at 229 amu]. Bottom panel: LC-DAD chromatogram (UV at 205 nm).

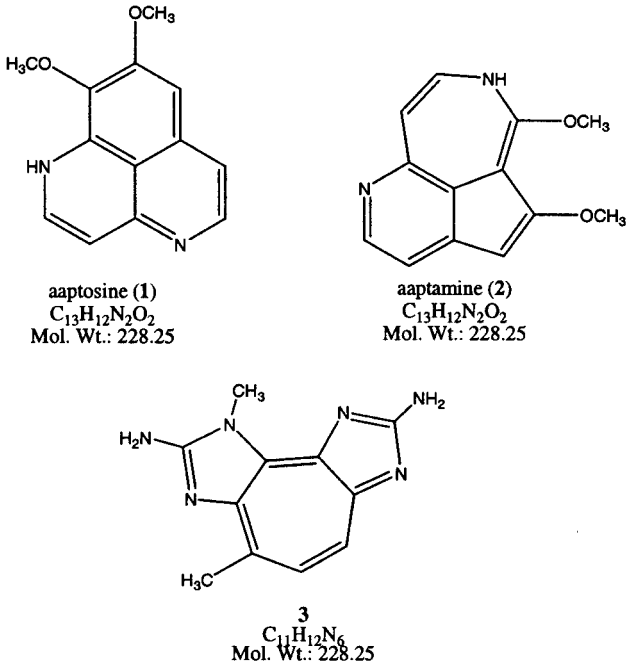

Figure 2 Possible structures for the enzyme inhibitory component of the crude extract of *Aaptos* sp. determined by LC-MS data and database searching.

injection of the crude sponge extract was made for the LC-NMR stopped-flow experiment.

Results

A HTS assay was designed to identify inhibitors of an enzyme which could be developed into potential drug candidates. HTS of Monsanto's collection of natural products extracts identified 283 samples which exhibited inhibitory activity against this enzyme. A combination of biological and chemical prioritization schemes was used to select extracts for further characterization.

The crude dichloromethane extract of the sponge *Aaptos* sp. collected in Manado, Indonesia was identified by the HTS effort for this enzyme (IC_{50}=18 μg/ml). In order to identify the active component in this crude extract, it was analyzed by LC-MS and LC-DAD methods. Fractions of the HPLC eluent were collected during the LC-MS experiment and assayed for enzyme activity to determine which peak in the chromatogram possessed the observed activity. This effort allowed the peak at 10.3 min to be identified as the single, active component in the crude extract (Figure 1). The mass spectral data for this peak indicated a molecular weight of 228 amu and the DAD data provided UV maxima of 237, 255, 309, and 375 for this compound. These parameters were used to search the marine natural products database, Marinlit [3]. Three compounds fit this data set (Figure 2) — aaptamine (Figure 2, structure 1) [4], aaptosine (Figure 2, structure 2) [6], and the paragracine relative (Figure 2, structure 3) [2]. It was clear that the data at hand did not allow us to confidently discern these three compounds and that more information would be required to make a definitive identification of the active component in the sponge extract.

In order to obtain more structural information about the active component in the sponge extract, a stopped-flow LC-NMR experiment was conducted on the crude extract using the same HPLC column and conditions as used for the LC-MS and LC-DAD experiments. The only difference in the chromatography conditions was the use of D_2O as the aqueous component of the HPLC solvent system. The eluent from the HPLC column was directed through a UV detector and then through the LC-NMR flow probe. The volume between the two detectors was calibrated and the solvent flow was stopped after the appropriate delay following observation of the UV peak previously observed at 10.3 min in the LC-MS experiment. This placed the peak of interest in the NMR active region of the LC-NMR flow probe. The ^{1}H NMR spectrum of this peak was acquired using a WET solvent suppression routine [8] to eliminate the signals due to the acetonitrile, its ^{13}C satellites, and water in the HPLC eluent. The spectrum obtained (Figure 3) revealed four coupled NMR signals and one singlet in the aromatic region (6–8 ppm) and two large singlets at 3.83 and 3.97 ppm. This information could be used to eliminate the paragracine relative (Figure 2, structure 3) from consideration because only two aromatic protons would have been observed for this molecule and the NH protons would have exchanged with the deuterium oxide in the HPLC solvent. ^{1}H NMR values from the literature for aaptamine (Figure 2, structure 1) [4] and aaptosine (Figure 2, structure 2) [6] were compared to the experimental values obtained by LC-NMR (Table 1). The NMR data for aaptamine matched closely (average $\Delta\delta$=0.14 ppm) with the experimental results while there were significant discrepancies between aaptosine and the LC-NMR data (average $\Delta\delta$=0.70 ppm). These data allow the confident assignment of the structure of

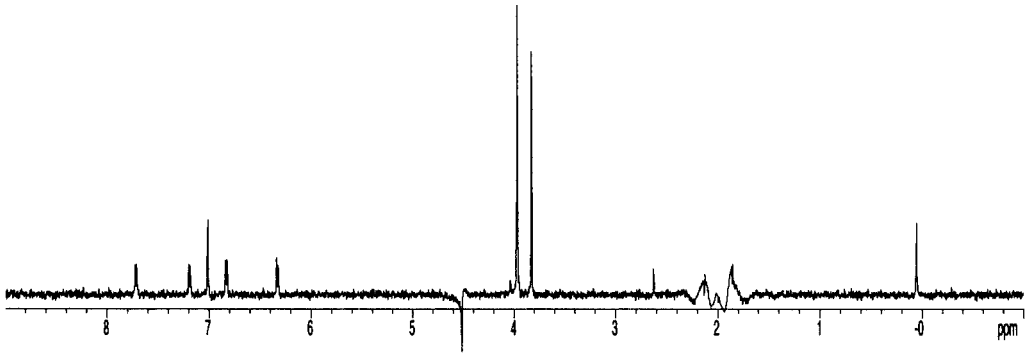

Figure 3 ^{1}H NMR (500 MHz) spectrum of the peak at 10.3 min in the crude extract of *Aaptos* sp. Data were acquired by stopped-flow LC-NMR using WET solvent suppression [4]. Disturbances in the baseline at 4.5 and 2.0 ppm are due to signals from residual H_2O and acetonitrile.

Table 1 Comparison of ^{1}H NMR data from the stopped-flow LC-NMR experiment on the peak at 10.3 min in the crude extract of the sponge *Aaptos* sp., aaptamine [4], and aaptosine [6]

Experimental value[a]	Aaptamine[b]	Aaptosine[b]
7.72	7.90 (+0.18)	8.97 (+1.25)
7.19	7.45 (+0.26)	8.95 (+1.76)
7.01	7.18 (+0.17)	7.92 (+0.91)
6.83	6.93 (+0.10)	7.15 (+0.33)
6.33	6.52 (+0.19)	6.14 (−0.19)
3.97	4.03 (+0.06)	3.97 (0.00)
3.83	3.86 (+0.03)	3.40 (−0.43)

[a]Values in parts per million (500 MHz). Solvent conditions approximately 43% ACN in D_2O (0.01% TFA).
[b]Numbers in parentheses represent the difference in chemical shifts of comparable signals between experimental and literature values.

aaptamine to the active component of the sponge *Aaptos* sp. Purified aaptosine exhibited an IC_{50}=120 μM when assayed against the isolated enzyme.

Discussion

The dichloromethane extract of the marine sponge *Aaptos* sp. inhibited the activity of the enzyme screened in our HTS program. The marine alkaloid aaptamine (Figure 2, structure 1) was determined to be responsible for the inhibitory activity using a combination of tandem analytical techniques. LC-MS was used to determine that the molecular weight of the active constituent was 228 amu and LC-NMR provided the necessary structural information to confidently assign the structure of aaptamine. The combination of these two tandem analytical techniques provides an example of how LC-NMR can be incorporated into a natural products drug discovery program to increase the efficiency of the dereplication process.

At this time, the limited sensitivity of LC-NMR will prevent it from being as widely used a tandem analytical technique as LC-MS is today, but use of stopped-flow measurements can increase the sensitivity down into the nanogram range. NMR spectral data obtained using LC-NMR provide structural information, which other methods cannot. This provides a useful complement to the more sensitive and higher throughput methods of LC-MS and LC-DAD. Due to the limits in sensitivity and the lack of searchable NMR databases, LC-NMR is not yet ready for use as a frontline dereplication technique. But as we have demonstrated here, it can be used to augment data obtained using more traditional dereplication techniques such as LC-MS and LC-DAD to allow for an efficient dereplication strategy. We recommend that LC-MS and/or LC-DAD methods continue to be used in their current roles for dereplication. The majority of samples can be confidently identified by these methods. But in the instances where the LC-MS and LC-DAD data provide numerous answers or when structural isomers are encountered, LC-NMR may provide the necessary data to complete the identification.

Acknowledgements

The authors acknowledge Dr. Sohrab Habibi-Goudarzi and Mr. Lawrence Goodwin for assistance in obtaining the LC-MS data and Drs. John Likos and Janice Gard for assistance in obtaining the LC-NMR data. Drs. Mark Hamann and Marc Slattery (University of Mississippi) are acknowledged for their assistance in the collection of samples from Indonesia.

References

1 Corley DG and RC Durley. 1994. Strategies for database dereplication of natural products. *J Nat Prod* 57: 1484–1490.
2 Komoda Y, M Shimizu and M Ishikawa. 1984. Structures of biologically active minor bases related to paragracine from *Parazoanthus gracilis*. *Chem Pharm Bull* 32: 3873–3879.
3 Marinlit — a database on marine natural product literature. John W. Blunt and Murray H. Munro, University of Canterbury, Christchurch, New Zealand.
4 Nakamura H, H Wu, R Abe, J Kobayashi, Y Ohizumi and Y Hirata. 1983. Structures of physiologically active substances isolated from the Okinawan marines sponges. *Tennen Yuki Kagobutsu Toronkai Koen Yoshishu* 26: 118–125.
5 Pannell LK and N Shigematsu. 1998. Increased speed and accuracy of structural determination of biologically active natural products using LC-MS. *Am Lab* 30 (7): 28–30.
6 Rudi A and Y Kashman. 1993. Aaptosine — a new cytotoxic 5,8-diazabenz[*c*,*d*]azulene alkaloid from the Red Sea Sponge *Aaptos* sp. *Tetrahedron Lett* 34: 4683–4684.
7 Shigematsu N. 1997. Dereplication of natural products using LC/MS. *J Mass Spectrom Soc Jpn* 45: 295–300.
8 Smallcombe SH, SL Patt and PA Keifer. 1995. WET solvent suppression and its application to LC-NMR and high-resolution NMR spectroscopy. *J Magn Reson A* 117: 295–303.