npg

# REVIEW ARTICLE

# Dereplication of microbial extracts and related analytical technologies

Tatsuya Ito and Miyako Masubuchi

**Natural products still continue to have an important role as a resource of various biologically active substances. Dereplication is a key process in natural product screening that analyzes the extracts of microbial fermentation broths or plant samples. In this review article, we describe and discuss the analytical techniques of dereplication and related technologies in the following sections: 1. Direct detection from microbial colonies. 2. Ultra high performance liquid chromatography (UHPLC)-MS profiling for library construction. 3. Micro-fractionation to identify active peaks. 4. Quantification of small-amount compounds. 5. Structure identification from small amounts. Using these techniques, the desired compound in the mixture library can be rapidly identified.**

## INTRODUCTION

Natural products and their derivatives continue to have an important role as a resource for various pharmaceuticals and biologically active substances. Some synthesized drugs were inspired by the scaffold of natural products and, even now, natural products are undoubtedly useful for discovering pharmaceutical leads.[1–5] However, after the great strides made in high-throughput screening (HTS) and synthetic compound libraries, natural product screening has been withdrawn in some pharmaceutical companies because it requires special know-how, and identifying active compounds takes time and is laborious.[6] These drawbacks mean that, to make natural product screening feasible, it is essential to develop an optimized assay system for an optimized natural product library and to use high-level analytical techniques for dereplication.

Screening procedures that explore a natural product as a source for a lead are roughly classified into two methods: either the screening is carried out using previously purified natural compounds, or the crude extract mixture is used in a primary screening, and after the active principle has been detected, it is purified by activity-guided purification. For the former, there are plenty of commercially purified natural product libraries that can be purchased to expand the screening quality to the same rank as a synthetic compound library. However, commercially available natural products cannot provide the wide diversity and wide area of chemical space that are sought for natural products. Alternatively, building a purified natural products library by purifying the compounds ourselves would require thousands of man-hours to construct a library sufficient in number and diversity.

On the other hand, because each crude extract contains a number of compounds, constructing a crude extracts library is not as laborious as making a purified compounds library and the range of diversity may be comparable with a large systematic compound library. However, a sophisticated and robust screening system with high specificity and sensitivity is needed to detect the target compound in the various compound mixtures.[7,8] In addition, dereplication strategy is required to eliminate known and redundant compounds, indicate which peak should be efficiently purified and accumulate knowledge for structure determination. Further, if the natural product screening is synchronized with pure compound HTS, the isolation and structure elucidation must be made time competitive.

The word 'dereplication' was used in the first CRC Handbook of Antibiotic Compounds published in 1980 to refer to recognizing and eliminating the active substances already studied in the early stage of the screening process.[9,10] Thereafter, it has been considered a stage subsequent to the preliminary screening in the process of discovering new pharmacologically active substances in a mixture of natural products. A systematic dereplication method for crude extract samples that can be combined with HTS has been developed using LC-UV-(MS).[11,12] Dereplication was carried out in combination with microbiological study, such as identifying producer microbes by morphology and gene analysis, and sometimes used comprehensively in combination with various informatics techniques.[13] Currently, the term dereplication is used in a broader sense in the screening of natural products mixtures, as shown in Figure 1. In this manuscript, we discuss dereplication and its related analytical techniques in the following five sections.
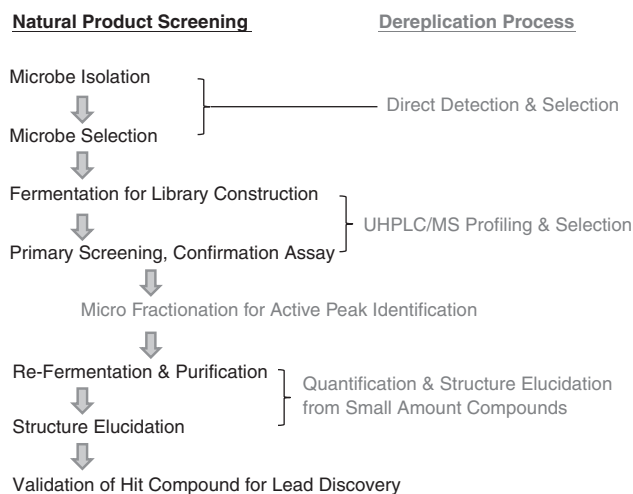
Research Division, Chugai Pharmaceutical Co., Ltd., Kamakura, Kanagawa, Japan
Correspondence: Dr T Ito, Research Division, Chugai Pharmaceutical Co., Ltd., 200 Kajiwara, Kamakura, Kanagawa 247-8530, Japan.
E-mail: itotty@chugai-pharm.co.jp

**Natural Product Screening**          Dereplication Process

Microbe Isolation

Microbe Selection ──── Direct Detection & Selection

Fermentation for Library Construction

Primary Screening, Confirmation Assay ──── UHPLC/MS Profiling & Selection

Micro Fractionation for Active Peak Identification

Re-Fermentation & Purification ──── Quantification & Structure Elucidation from Small Amount Compounds

Structure Elucidation

Validation of Hit Compound for Lead Discovery

**Figure 1** Dereplication process in the natural products screening. A full color version of this figure is available at *The Journal of Antibiotics* journal online.

1. Direct Detection from Microbial Colonies
2. UHPLC-MS Profiling for Library Construction
3. Micro-fractionation for Active Peak Identification
4. Quantification of Small-Amount Compounds
5. Structure Identification from Small Amounts

In this review, the dereplication process and related technologies are summarized from the aspect of analytical chemistry.

## DIRECT DETECTION FROM MICROBIAL COLONIES

If compounds such as mycotoxins or other nuisance compounds can be detected directly from the surface of a microbial colony that has been newly obtained, it is possible to avoid culturing unnecessarily. Alternatively, some unrecognized colonies may be useful, if they have a spectrum never seen before. MS is an effective way to directly analyze natural product samples. The recent development of ambient MS technology means that an untreated sample can be directly analyzed at atmospheric pressure,[14] which prevents the need for separation by LC, and analysis can be completed in an extremely short time.

Desorption electrospray ionization (DESI) is an ambient ionization technique that uses a high voltage ion source for the electrospray to make an ionizing liquid mist. The charged liquid mist is sprayed onto the sample surface so that surface molecules are ionized in atmospheric pressure before being moved into the MS analyzer.[14,15] A modified DESI method, nano-DESI, uses two capillary systems at the ion source; the first capillary is connected to a nano-LC pump and charged at high voltage, and the second capillary is connected to an MS inlet. The joint between the first and second capillaries, termed as liquid bridge, is placed slightly above the surface of the microbial colony.[16] Ionization occurs over a small area of the colony and the metabolite ions are detected. Watrous *et al.*[17] analyzed live microbial colonies of *Bacillus subtilis*, *Streptomyces coelicolor*, *Mycobacterium smegmatis* and *Pseudomonas aeruginosa* using a nano-DESI-MS/MS instrument and monitored the metabolic profile. They demonstrated the microbial molecular networking and interference of live colonies.

The nano-DESI method does not seem to be easy to use in microbial laboratories generally because of capillary clogging. To directly analyze microbial colonies, Hsu *et al.*[18] employed a liquid microjunction surface sampling probe (LMJ-SSP), which is a more robust direct ionization method based on the principle of electrospray ionization (ESI) that uses a coaxial tube geometry sampling probe.[19] They analyzed yeast, filamentous fungi, pathogens and marine-derived bacteria collected worldwide by this method.

Liquid extraction surface analysis, automated chip-based nanoelectrospray ionization technique, was used for direct analysis of live bacterial colonies on the culture media. Kai *et al.*[20] demonstrated that the antibiotics such as thiazolyl peptide could be correctly identified by the combination of liquid extraction surface analysis and Orbitrap.

Another atmospheric pressure ionization method is direct analysis in real time (DART), which uses the plasma-based technique. The ionization process involves an interaction between the analyte molecule and a gas ($N_2$, Ne or He) that has been electrically excited to interact directly with the sample surface and ionize the molecule. This ion source has already been commercialized by several companies and can be used with some mass detectors. Direct analysis of microbial colonies by DART-MS has not yet been reported; although Zachariasova *et al.*[21] tried to analyze contaminated mycotoxins directly from beer, it seems that DART-MS is not as sensitive in detecting mycotoxins as ultra high performance liquid chromatography (UHPLC)-MS. Vaclavik *et al.*[22] analyzed 24 different mycotoxins spiked in grain using a DART-Orbitrap-MS and found that zearalenone and deoxynivalenol produced by *Fusarium* sp. could be analyzed quantitatively, but ions of T-2 toxin, HT-2 toxin and ochratoxin A could not be observed.

It is expected that matrix assisted laser desorption ionization (MALDI)-time of flight (TOF)-MS will be used for the same purposes as above, and the use of MS imaging is also expanding.[23,24]

At present, there are only a few reports on directly detecting metabolites from a microbial colony by ambient MS technology, but in the future this technology could be utilized as a means of identifying substances quickly. The time may come when MS detectors installed beside a clean bench can be used to detect and isolate microorganisms in a culture.

## UHPLC-MS PROFILING FOR LIBRARY CONSTRUCTION

When screening libraries are constructed from microbial culture extracts, a lot of strains are cultured in small bottles or flasks, extracted with the appropriate organic solvent, and the extracts are stored in storage tubes before being dispensed into 96-well microtiter plates, and from there into 384- and/or 1536-well plates. At this point, unlike a synthesized compound library, it is not known what compounds are included in the extract library or what the mixture profiles are. If the same method of screening is used as in a synthesized compound library, the same well-known compounds would be given at the top of the hit lists. Alternatively, if the samples are randomly analyzed by HPLC or LC-MS, a considerable number of crude extract samples would be found to contain almost no lead-like compounds.

Because of this situation, we realized that samples needed to be selected by a high-throughput method of analysis before constructing an extract screening library. This can be efficiently accomplished with an HPLC connected to MS and photo-diode array detectors, so that the system can analyze and identify many substances in one injection and also carry out profiling of the mixture. In 2004, UHPLC was commercialized using a column packed with chemically modified silica particles (C18, C8, Phenyl, etc.), whose diameters were less than 2 μm, and now this system is widely used.[25] Because the detectors can be connected to the UHPLC, the system can obtain a large number of data points quickly, and examples of analysis are increasing.[26–31]

In particular, UHPLC can analyze at a high resolution in a short time compared with HPLC, making it possible to analyze a lot of samples.

Various detectors can be connected to the UHPLC or HPLC but none are completely comprehensive. The general ionization methods of LC-MS are ESI, atmospheric pressure chemical ionization and atmospheric pressure photo ionization, each of which has both positive and negative ion modes, although it depends on the compound whether it should be detected with both ion modes or with just one.[32] Sample analysis with both the positive and negative of ESI means that the detectable range is wide. Although many of the quadrupole-type low resolution (LR)-MS methods can acquire both polarities simultaneously by rapidly switching polarity, most HR-MS, such as TOF and Orbitrap, cannot switch polarity rapidly, so samples must be analyzed twice to gain data on both polarities. To analyze the data, metabolomics technology can be applied and multivariate analysis is a well-established procedure for LC-MS data.[33–36] Cluster analysis of LC-MS data by several types of freeware or commercial software has been reported.

We have developed a high-throughput profiling method that can analyze within 3.1 min per cycle using an instrument that combines the single quadrupole MS with UHPLC.[37] Because data processing takes much time, we used LR-MS, but a higher level of identification accuracy is possible with HR-MS. The analysis was conducted using a C18 reverse phase column, and extracts of liquid culture broth (by butanol) or of solid cultures (by methanol) were analyzed by both positive and negative ion modes. Peak picking from the raw data was carried out by ACD MS Manager with IntelliXtract (the latest version is ACD Spectrus MS workbook suite V2012). Then the number of occurrences of peaks in all 16 000 samples was calculated by the peak identifier, using a combination of MS (nominal $m/z$) and retention time (RT, min). The profile of the crude extract components was visualized in TIBCO Spotfire software to produce a scatter plot that reveals what substances are in the sample, what their MW is and how frequently the substance occurs in the library (Figure 2). The combination of MS and RT data can help us to identify some substances.

Another database (DB) for managing fraction libraries that combines MS and RT is established by Kato et al.[38] and called NP Plot. This system uses not only MS data but also a UV spectrum to identify novel metabolites. Combining RT and MS has also been attempted in medicinal plant metabolite profiling.[39] Bertrand et al.[29] carried out profiling using the UHPLC-TOF-MS. They tried various solvents to make the extraction from solid culture and selected one (a ratio of dichloromethane:methanol:water at 64:36:8) that could dissolve both lipid and hydrophilic compounds and was therefore appropriate for profiling diverse microbial metabolites. The LC-MS data were first converted to nCDF, a common data format, then to a peak list using MZmine Software and then to a peak table. By this method, they analyzed the changes of metabolites in the co-culture of the fungi.

When UHPLC-MS analysis was first used, the scan speed of MS could not keep up with the high resolution of UHPLC, so the resolution of UHPLC was lowered on purpose. Currently, MS analysis at a sufficiently high speed is possible if we analyze the positive and
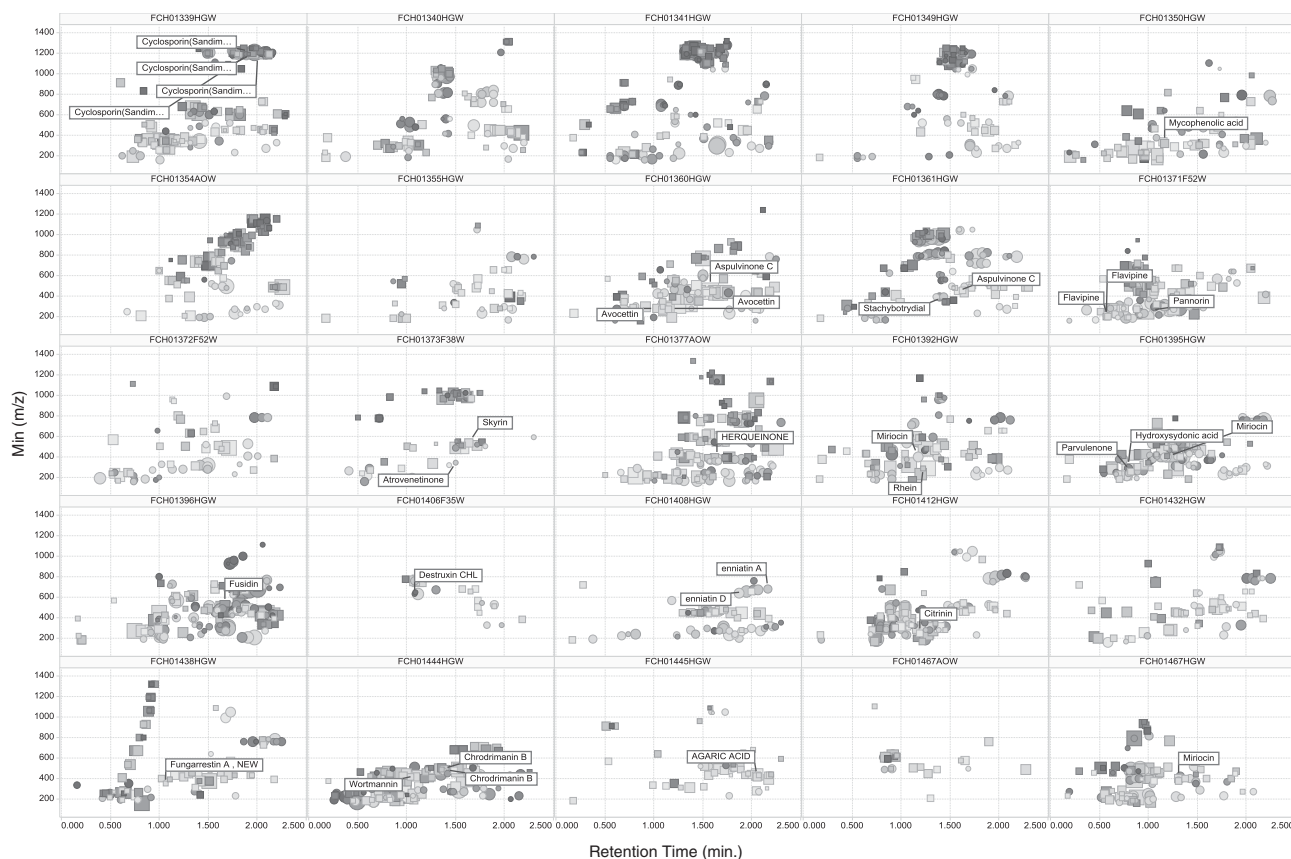


**Figure 2** UHPLC-MS profiling of microbial extracts. Horizontal axis is retention time (min); vertical axis is mass ($m/z$). Each cell shows one sample. Circles indicate positive ions, and squares indicate negative ions of mass. Symbol size represents peak intensity. Frequency of appearance throughout library was indicated by color. A full color version of this figure is available at *The Journal of Antibiotics* journal online.

negative ion modes separately, and the rate-limiting factor would be data processing, because analyzing large amounts of samples at high speed and at high resolution gives large amounts of data. However, this will be solved in the near future by the rapid development of informatics technology. Growing the number of library samples increases the time taken from isolating the microorganism culture to profiling the samples, so a server system that processes the analysis data automatically is desirable.

## MICRO-FRACTIONATION FOR ACTIVE PEAK IDENTIFICATION

Active peaks cannot be identified merely by looking at the data of LC-UV-MS or LC-UV. Ways of detecting biological activity using on-line systems that directly connect to HPLC have been reported previously,[40,41] but the range of assay systems they accept is limited. To identify an active component, a mixture sample needs to be fractionated, which consists of connecting a microscale fraction collector to an LC-UV-MS system and subjecting the collected fractions to various assays to reveal which peak is the active

component (Figure 3). Hook et al.,[11,12] researchers at Bristol Myers Company, described their micro-fractionation process. They connected the fraction collector to HPLC-UV or HPLC-UV-MS and used microtiter plates to collect fractions that were then subjected to HTS using robotic systems. This procedure has been used widely[42–44] and is called dereplication, in the narrow meaning of the word. Wagenaar[45] also reported a method of constructing a fraction library that was prepared before screening.

Of the equipment currently available for micro-fractionation, the C18 reverse phase column is the most versatile of HPLC columns. For special analysis, columns such as HILLIC,[46] CN, Amino and C4 are used, but they are not the first choice for analyzing unknown samples. Considering the amount of fractions to be collected, a column with a diameter of 3.0 or 4.6 mm (flow rate 1–3 ml min$^{-1}$) is recommended, and a column length of 50–150 mm is convenient. For a fractionation method using a larger diameter column, Bohni et al.[47] have fractionated using a 10 × 250 mm column to obtain fractions in an in vivo evaluation using Zebrafish. To connect a fraction collector and
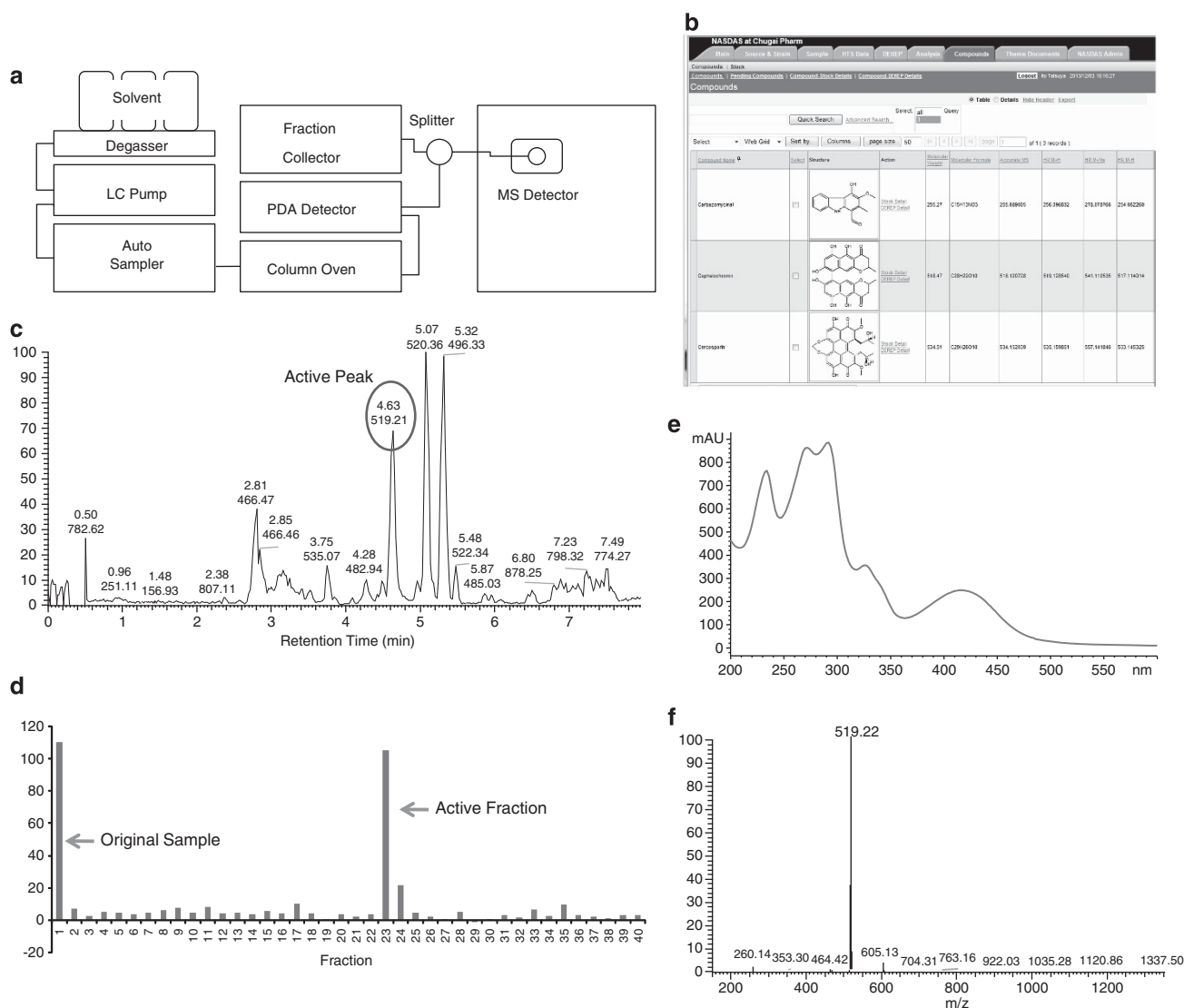


Figure 3 Micro-fractionation (dereplication) to identify active peaks. (a) Schematic diagram of LC-MS for dereplication. (b) View of the in-house DB (NASDAS) for natural products screening. (c) Total ion chromatogram of mixture sample by MS. (d) Activity of fractions. (e) UV spectrum of an active peak. (f) A mass spectrum (ESI +). The active compound was identified as cephalochromine from the mass and UV spectra. A full color version of this figure is available at The Journal of Antibiotics journal online.

MS at the same time, a splitter and a method of optimizing the splitting ratio are required. To remove the solvent from fractions, a centrifugal evaporator is mainly used but this has two major drawbacks: the solvents used in reverse phase HPLC are water rich, which makes this step time-consuming and a bottleneck of the micro-fraction assay; second, unstable compounds may easily degrade during the evaporation. In order to avoid evaporation step, trap column (solid phase extraction cartridge) can be used to concentrate the fraction like LC-SPE-NMR technique.[44] The unstable compounds might be captured by this method.

As mentioned in former chapter, the ion sources currently used in LC-MS instruments, ESI, atmospheric pressure chemical ionization and atmospheric pressure photo ionization, have been commercialized. For natural products, ESI is the most widely used, and a large number of reports have been published since the 1990s.[31] There is no doubt that ESI is the ion source that can detect most things as exhaustively as possible. Also, because analysis of both positive and negative polarities simultaneously is appropriate for comprehensive detection, TFA is widely used as a pH adjuster in LC-MS because it has low UV absorption and a low boiling point, but unfortunately it cannot be used in negative ion mode because it lowers the sensitivity.

As MS instruments, the quadrupole-type system, ion trap, TOF and Orbitrap can be used for dereplication; however, TOF or Orbitrap are preferable because they are capable of HR-MS measurement. Using MS data, it is possible to search compounds from a commercially available natural products DB. We usually use Antibase (http://wwwuser.gwdg.de/~hlaatsc/antibase.htm), or Chapman & Hall Dictionary of Natural Products (CHDNP) (http://dnp.chemnetbase.com/intro/). In Antibase 2013, 41 864 metabolites derived mainly from microorganisms are listed and updated once a year; in CHDNP, all natural products are covered and updated twice a year.

Generally, it is difficult to identify compounds only by MS data, and referring to the UV spectra in addition to MS data can be helpful because a characteristic UV spectrum is easy to identify. In some DB records, such as Antibase and CHDNP, the value of UV $\lambda_{max}$ wavelength is described but not in all of them. RT also helps to identify the compounds and, although it is not possible to find RT in a commercial DB, it is possible to refer to data in published papers. Nielsen et al.[48,49] summarized the MS, UV and RT data of 475 fungal metabolites, and these data are useful for identifying compounds. Boswell et al.[50] described that MS data and RT are orthogonal, so compounds can be identified more efficiently from a combination of RT and LR-MS than with only HR-MS.

MS/MS spectra can be easily obtained during micro-fractionation with LC-UV-MS analysis. Ion trap, triple quadrupole and Q-TOF-MS can acquire MS/MS data, and even single-stage quadrupole and TOF apparatus can also get MS/MS spectra using in-source collision-induced decay. Elucidating a structure by MS/MS spectra has been reported,[51,52] but it is not easy in natural products that have a complex scaffold of compounds. There is a public DB of electron impact-MS spectra, but none for the ESI-MS/MS spectrum. Recently, the public DB Mass Bank, in which 30 000 MS/MS spectra have been collected, can be used on a web environment,[53] but the number of microbial natural compounds is small.

As explained earlier, UV spectra acquired by a photo-diode array detector in an HPLC or UHPLC system are important because, in contrast to MS spectra, it is easy to identify the compound group that has a characteristic UV chromophore, even in unknown derivatives. HPLC/UHPLC instruments are regulated by the software installed by the LC instrument maker, and can be classified into two classes: one is derived from the HPLC system (for example, Waters Empower) and one is from the MS system (for example, Waters MassLynx). The former (Empower, Agilent Chemstation, Thermo Dionex Chromerelon etc.) functions as a UV spectrum DB, which can store the whole spectrum of UV and is searchable making it convenient and easy to use. However, the latter system has no function for UV spectrum search in the DB, so if we want to use an MS instrument and a UV spectrum DB, we have to use two PCs for acquisition.

Accumulating in-house data on LC-UV-MS and on chemical structure is important for continuing the dereplication process, and some research groups have attempted to construct such a DB.[38,54,55] The purpose of constructing the DB is to identify newly analyzed data and to record library data on which compounds were identified from the samples. We have constructed an in-house DB called Natural Sample Data System (NASDAS) (Figure 3).

We have described dereplication with fractionation by reverse phase HPLC as being the most versatile, but we cannot deny that there are a few substances that cannot be captured as a peak. As a method that does not use the LC, there are solid phase extraction, ion exchange and a preparative TLC. In addition, Månsson et al.[56] have designed a dereplication system (E-SPE) using a type of SPE cartridge (LH20, SAX, Oasis MAX, SCX, etc.) and then analyzing the elution pattern of activity. High-throughput 96-well fractionation using the hydrophilic-lipophilic balance plate (Oasis HLB) was employed for the dereplication and efficient discovery of thiazolyl peptides.[57] The eluted fractions were analyzed by HR-FT-LC-MS. Normal phase chromatography is a powerful tool for analyzing compounds that are not suitable for reverse phase separation. It is also possible to observe the MS spectrum of the substance by DART-MS directly from TLC.[58]

The dereplication process involving micro-fractionation can be used not only to identify known substances but also to target peaks for purification, so that we only purify those in which a target peak has been identified. This is efficient compared with traditional purification that needs the fractionation-activity evaluation cycle to be repeated.

## QUANTIFICATION OF SMALL-AMOUNT COMPOUNDS
The amount of microbial metabolites that are produced is extremely diverse. Some are less than 1 mg in 1 l of culture, while others produce more than 1 g in 1 l. When crude extracts are screened biologically, the sample is counted as a hit whether a large amount of metabolite with weak activity or a trace amount of metabolite with strong activity is produced. It would be helpful to be able to quantitatively determine metabolites with a certain degree of accuracy in the dereplication step of the initial screening. On the other hand, compounds purified from a small amount of culture broth at the stage of initial purification are very small and difficult to weigh. If an authentic sample exists, a comparative quantification is possible by HPLC; if not, a method of absolute quantification must be used.

In the case of LC-UV, compounds with chromophores can be quantified with relatively high sensitivity but, depending on the compound, determining the quantity without a standard sample is difficult. When the compounds have end absorption, the amount can be guessed to a certain extent, but if compounds do not have any UV absorption at all, as in the case of lipids, some macrolides and sugars, it is difficult to determine the quantity by LC-UV. In LC-MS, the sensitivity is also affected by how easily a substance ionizes, as well as on the model of MS apparatus used (even when they have the same ESI), which again makes absolute quantification difficult.

A chemiluminescent nitrogen detector (CLND) is an HPLC detector that can quantify the amount of nitrogen with high accuracy

and very high sensitivity.[59] However, CLND cannot detect compounds that contain no nitrogen, and it is not possible to use nitrogen-containing solvents, such as acetonitrile.

The first weight detector connectable to HPLC to be marketed was an evaporative light-scattering detector (ELSD).[60] The principle of ELSD is that the sample is nebulized in the mobile phase of HPLC and, after excluding the large water droplets, solute is taken as solid particles by heating; finally, the solid particles of the sample are detected by the intensity of scattered light. A nano quantity analyte detector (NQAD)[61] is a detector similar to ELSD, but the sensitivity is greater because the size of solid particles is enhanced by water vapor. Unlike NMR and CLND, signals of ELSD and NQAD correspond to the weight rather than the number of moles of substance. The relationship between the response and the amount of substance gives a straight line in the Log-Log graph in ELSD.[62] The sensitivity of ELSD and NQAD is better than that of refractive index detectors.

Weight can also be detected by Corona's charged aerosol detection (CAD) system, which was developed soon after ELSD, but has sensitivity several times greater. The measurement principle is similar to ELSD in that it makes solid particles of the solute by nebulizing the mobile phase of HPLC. The solid particles are charged with a Corona electrode, and then quantifies the charged particles using like an electrochemical detector.[63] Similarly to ELSD, the signal corresponds to the analyte weight, and the relationship between the responses and the amount of substance is a straight line in the Log-Log graph. The response from the Corona CAD depends on the mobile phase, such as acetonitrile content, but if the composition of the mobile phase is maintained at a constant rate before entering the detector with another gradient pump, the so-called reverse gradient, the response reflects only the analyte weight. A second generation system, the Corona Ultra detector, has been developed and can be used with UHPLC. Hutchinson et al.[61] compared the ELSD, NQAD, Corona CAD and Corona Ultra detectors. They described NQAD as the most sensitive, and Corona CAD and Corona Ultra best in terms of reproducibility. All detectors are less sensitive with semi-volatile molecules (non-polar and MW < 270).

Quantification by NMR (qNMR) is a powerful tool for purified natural products that, similarly to CLND, can obtain the concentration in molar units by qNMR[64,65] and the peak area of $^1$H-NMR signal correlates to the number of protons. Recently, the linearity of the receiver gain of NMR instruments has been improved, making it unnecessary to use an internal or external standard for quantification.[66] Currently, a pulse sequence function, named electronic reference to access in vivo concentrations (ERETIC) that is included as default on NMR instruments can be used to quantify without standard chemicals.

In the flow of natural products screening, active compounds are quantified by Corona CAD/Ultra before purification, and qNMR is useful to quantify purified compounds.

## STRUCTURE IDENTIFICATION FROM SMALL AMOUNTS

It is possible to increase the efficiency of lead discovery if the structure is determined with as small amount of culture broth as possible. When the active peak is identified through micro-fractionation, and a small amount of the active compound has been isolated, the purified compound may be subjected to $^1$H-NMR acquisition as the first step. $^1$H-NMR is useful both for elucidating structure and for checking the impurity. Sometimes impurities that are invisible in LC-UV-MS can be seen in $^1$H-NMR, and in many cases $^1$H-NMR reveals whether the purification went well or not.

From microbial culture broth, many compounds can be purified by just combining solvent extraction and reversed phase preparative HPLC with gradient elution. In such a short-term purification, lipids are problematic impurities that are not detectable in LC-UV-MS analysis because they have low UV absorption, have difficulty of ionizing and high lipophilicity. In $^1$H-NMR data acquisition, we first recognize the lipid impurity by the large peak of its methylene chain. The contamination is adsorbed to the preparative column, and it is possible for the octadecyl group that comes off the stationary phase of the C18 column to be mixed in the purified samples. The latter may also become noticeable when a new preparative column is used in the mobile phase with a high acetonitrile ratio without washing thoroughly. To check how efficiently the column has been washed, monitoring with Corona CAD or ELSD is recommended. When an appropriate $^1$H-NMR spectrum has been obtained, we assign the peaks and elucidate the structure by the NMR with UV and MS data. Lang et al.[67] have developed a DB named Anti Marin that can display the number of aldehyde, singlet methyl or other singlet peaks that are easily identifiable in the $^1$H-NMR spectrum. Even for trace compounds, easily readable NMR data can be obtained from various types of LC-NMR and Cap NMR.[68] If the compound has some characteristic signals, it can be searched in the DB by entering the number of peaks.

The hyphenated technique of NMR with HPLC (LC-NMR) is a widely used technique that dispenses with the need to isolate the compound before NMR acquisition. The method is classified into two categories: one measures NMR directly on-line from LC using a flow cell, the other collects the peaks of HPLC in an SPE column and measures the eluent by heavy solvent. Some excellent reviews of LC-NMR have been published.[69,70]

In $^1$H-NMR of a mixture sample, clearly different values of signal integrals make it easy to distinguish the substances in the mixture, but when mixed substances show similar signal intensity with very different MWs, measuring by diffusion-ordered NMR spectroscopy (DOSY) spectrum is recommended. DOSY is a method of separating the mixed signals by utilizing the difference in the coefficient of molecular diffusion.[71] Recently, many NMR spectrometers capable of a magnetic field gradient have the pulse sequence installed by default, but DOSY struggles when diffusion coefficients are similar and/or when spectra overlap heavily. To overcome the problem posed by analog mixtures that show the same diffusion, a DOSY approach has been reported by Cassani et al.[72] for resolving flavonoid mixtures with a slow diffusion matrix (micellar sodium dodecyl sulfate).

Even if high quality $^1$H-NMR spectra are obtained, sometimes we cannot elucidate the structure. In that case, the various 2D-NMR spectra such as HSQC, $^1$H-detected HMBC, DQF-COSY, TOCSY, NOESY and ROESY are usually tried, and a planar structure of compound can be elucidated. Presently the NMR of 600 MHz equipped with a Cryo Probe (Cold Probe), which can measure two-dimensional NMR of the proton in sample amounts of 10 μg, is considered to show a high cost benefit in determining the structure of natural products. However, structural analysis is sometimes difficult even in simple molecules; if the compound has many hetero elements and quaternary carbons, the obtained information from NMR is limited even when the MW is small.

X-ray crystallography is an excellent method for determining the structure, including the stereochemistry of the molecule, but crystallization of the compound is often difficult and requires a certain amount of sample. Inokuma et al.[73] have developed an X-ray crystallographic analysis using crystalline sponge, which absorbs the compound solution into the tiny crystal of a porous complex that

they developed. As the guest molecules are oriented regularly in the pores, crystallographic analysis clearly determines the absorbed guest structure along with the host frameworks. A great advantage of this method is clear when the structure of trace amounts of samples is elucidated. For example, they have succeeded in elucidating the structure from a small amount of compound separated by analytical HPLC (LC-SCD). The crystalline sponge method is not applicable to all natural products, because it requires that the molecular size should be smaller than the pores (MW of 500 or less), and the molecular polarity be slightly lower, but in the future it may become one of the most powerful methods of analyzing the structure of trace amounts of natural products.

## CONCLUSION

This paper described various processes for screening natural products from microbial culture broth, that means identifying their active principle, and finally purifying the active compound or determining its structure. Of these processes, we focused on dereplication from the aspect of analytical chemistry.

Direct detection from the microbial colony using the ambient MS technology might be used widely. MS instruments are gradually becoming miniaturized and using these compact instruments may make identifying a substance as easy as picking the microbe from petri dish. It may be possible to identify metabolites at the same time as observing the morphology of microorganisms, so that only beneficial microorganisms are cultured.

Advances in UHPLC-MS technology have improved methods of profiling rapidly in the last 10 years. If profiles of the component are analyzed before building a mixture library, samples containing frequent hitter compounds[74] or samples without any metabolites can be excluded, and as a result we can screen only useful broth. In addition, it is possible to construct a focused library in advance, because LC-MS data of a substance can predict the polarity from the RT and the MW from the MS. For example, it would be easy to build a library including metabolites with a MW of more than 600 for a protein–protein interaction (PPI) inhibitor that requires a certain molecular size.[8]

In dereplication by micro-fractionation, the active peaks and then the substance can be identified from the UV and MS. Dereplication can also indicate the appropriate purification condition for a preparative LC method.

Absolute quantification of the amount of active components in the early stage of purification is valuable for lead discovery by quantitatively analyzing the activity. LC-Corona CAD (or Corona Ultra) and ERETIC of $^1$H-NMR can quantify on a scale of 1–10 μg, even though it is impossible to actually weigh such small amounts.

Recently, X-ray crystallographic analysis technology of trace natural products, using a crystalline sponge has been developed and, in the near future, the concept of structure determination of natural products may change.

In this paper, we have presented natural product screening in terms of analytical chemistry, but constructing an appropriate assay system is the most important aspect when screening a mixture. Alternatively, there is a lot of room for considering how to fit the natural products library to the assay system; for example, how much pre-processing is needed. It is difficult to obtain a lead compound by simply applying the HTS method for a pure chemical compound to the natural product mixture library. The important thing is how to construct a mixture library that contains the diverse active compounds with an optimized assay system.

1 Demain, A. L. & Sanchez, S. Microbial drug discovery: 80 years of progress. *J. Antibiot.* **62,** 5–16 (2009).
2 Liu, X. *et al.* Bioprospecting microbial natural product libraries from the marine environment for drug discovery. *J. Antibiot.* **63,** 415–422 (2010).
3 Ōmura, S. Microbial metabolites: 45 years of wandering, wondering and discovering. *Tetrahedron* **67,** 6420–6459 (2011).
4 Newman, D. J. & Cragg, G. M. Natural products as sources of new drugs over the 30 years from 1981 to 2010. *J. Nat. Prod.* **75,** 311–335 (2012).
5 Cragg, G. M. & Newman, D. J. Natural products: a continuing source of novel drug leads. *Biochim. Biophys. Acta* **1830,** 3670–3695 (2013).
6 Bérdy, J. Thoughts and facts about antibiotics: where we are now and where we are heading. *J. Antibiot.* **65,** 385–395 (2012).
7 Henrich, C. J. & Beutler, J. Matching the power of high throughput screening to the chemical diversity of natural products. *Nat. Prod. Rep.* **30,** 1284–1298 (2013).
8 Takagi, M. & Shin-Ya, K. Construction of a natural product library containing secondary metabolites produced by actinomycetes. *J. Antibiot.* **65,** 443–447 (2012).
9 Langlykke, A. *CRC Handbook of Antibiotic Compounds (Bardy, J. et al.) (CRC, Boca Raton, FL, 1980).*
10 Vanmiddlesworth, F. & Cannell, R. J. P. Dereplication and partial identification of natural products. *Nat. Prod. Isolation* **4,** 279–327 (1998).
11 Hook, D. J., More, C. F., Yacoucci, J. J., Dubay, G. & O'Connor, S. Integrated biological-physicochemical system for the identification of antitumor compounds in fermentation broths. *J. Chromatogr.* **385,** 99–108 (1987).
12 Hook, D. J. Approaches to automating the dereplication of bioactive natural products–the key step in high throughput screening of bioactive materials from natural sources. *J. Biomol. Screen.* **2,** 145–152 (1997).
13 Ibrahim, A., Yang, L. & Johnston, C. Dereplicating nonribosomal peptides using an informatic search algorithm for natural products (iSNAP) discovery. *Proc. Natl Acad. Sci. USA* **109,** 19196–19210 (2012).
14 Harris, G. A., Galhena, A. S. & Fernández, F. M. Ambient sampling/ionization mass spectrometry: applications and current trends. *Anal. Chem.* **83,** 4508–4538 (2011).
15 Monge, M. E., Harris, G. A., Dwivedi, P. & Fernández, F. M. Mass spectrometry: recent advances in direct open air surface sampling/ionization. *Chem. Rev.* **113,** 2269–2308 (2013).
16 Roach, P. J., Laskin, J. & Laskin, A. Nanospray desorption electrospray ionization: an ambient method for liquid-extraction surface sampling in mass spectrometry. *Analyst* **135,** 2233–2236 (2010).
17 Watrous, J. *et al.* Mass spectral molecular networking of living microbial colonies. *Proc. Natl Acad. Sci. USA* **109,** E1743–E1752 (2012).
18 Hsu, C.-C. *et al.* Real-time metabolomics on living microorganisms using ambient electrospray ionization flow-probe. *Anal. Chem.* **85,** 7014–7018 (2013).
19 Van Berkel, G. J., Kertesz, V. & King, R. C. High-throughput mode liquid microjunction surface sampling probe. *Anal. Chem.* **81,** 7096–7101 (2009).
20 Kai, M., González, I., Genilloud, O., Singh, S. B. & Svatoš, A. Direct mass spectrometric screening of antibiotics from bacterial surfaces using liquid extraction surface analysis. *Rapid Commun. Mass Spectrom.* **26,** 2477–2482 (2012).
21 Zachariasova, M. *et al.* Analysis of multiple mycotoxins in beer employing (ultra) -high-resolution mass spectrometry. *Rapid Commun. Mass Spectrom.* **24,** 3357–3367 (2010).
22 Vaclavik, L., Zachariasova, M., Hrbek, V. & Hajslova, J. Analysis of multiple mycotoxins in cereals under ambient conditions using direct analysis in real time (DART) ionization coupled to high resolution mass spectrometry. *Talanta* **82,** 1950–1957 (2010).
23 Watrous, J. D. & Dorrestein, P. C. Imaging mass spectrometry in microbiology. *Nat. Rev. Microbiol.* **9,** 683–694 (2011).
24 Kersten, R. D. *et al.* A mass spectrometry-guided genome mining approach for natural product peptidogenomics. *Nat. Chem. Biol.* **7,** 794–802 (2011).
25 Mazzeo, J. R., Neue, U. D., Kele, M., Plumb, R. S. & Corp, W. A new separation technique takes advantage of sub-2-μm HPLC. *Anal. Chem.* **77,** 460A–467A (2005).
26 Chan, E., Yap, S. & Lau, A. Ultra-performance liquid chromatography/time-of-flight mass spectrometry based metabolomics of raw and steamed Panax notoginseng. *Mass Spectrom.* **21,** 519–528 (2007).
27 Dan, M. *et al.* Metabolite profiling of Panax notoginseng using UPLC-ESI-MS. *Phytochemistry* **69,** 2237–2244 (2008).
28 Tchoumtchoua, J., Njamen, D., Mbanya, J. C., Skaltsounis, A.-L. & Halabalaki, M. Structure-oriented UHPLC-LTQ Orbitrap-based approach as a dereplication strategy for the identification of isoflavonoids from Amphimas pterocarpoides crude extract. *J. Mass Spectrom.* **48,** 561–575 (2013).
29 Bertrand, S. *et al.* Detection of metabolite induction in fungal co-cultures on solid media by high-throughput differential ultra-high pressure liquid chromatography-time-of-flight mass spectrometry fingerprinting. *J. Chromatogr. A* **1292,** 219–228 (2013).
30 Wolfender, J.-L. HPLC in natural product analysis: the detection issue. *Planta Med.* **75,** 719–734 (2009).

31 Wolfender, J., Marti, G. & Queiroz, E. F. Advances in techniques for profiling crude extracts and for the rapid identification of natural products: Dereplication, Quality control and metabolomics. *Curr. Org. Chem.* 1808–1832 (2010).

32 Nordström, A., Want, E., Northen, T., Lehtiö, J. & Siuzdak, G. Multiple ionization mass spectrometry strategy used to reveal the complexity of metabolomics. *Anal. Chem.* **80,** 421–429 (2008).

33 Moco, S. *et al.* A liquid chromatography-mass spectrometry-based metabolome database for tomato. *Plant Physiol.* **141,** 1205–1218 (2006).

34 Wishart, D. S. Current progress in computational metabolomics. *Brief. Bioinform.* **8,** 279–293 (2007).

35 De Vos, R. C. H. *et al.* Untargeted large-scale plant metabolomics using liquid chromatography coupled to mass spectrometry. *Nat. Protoc.* **2,** 778–791 (2007).

36 Suits, F., Lepre, J., Du, P., Bischoff, R. & Horvatovich, P. Two-dimensional method for time aligning liquid chromatography-mass spectrometry data. *Anal. Chem.* **80,** 3095–3104 (2008).

37 Ito, T., Odake, T., Katoh, H., Yamaguchi, Y. & Aoki, M. High-throughput profiling of microbial extracts. *J. Nat. Prod.* **74,** 983–988 (2011).

38 Kato, N., Takahashi, S., Nogawa, T., Saito, T. & Osada, H. Construction of a microbial natural product library for chemical biology studies. *Curr. Opin. Chem. Biol.* **16,** 101–108 (2012).

39 Cuthbertson, D. J. *et al.* Accurate mass-time tag library for LC/MS-based metabolite profiling of medicinal plants. *Phytochemistry* **91,** 187–197 (2013).

40 Potterat, O. & Hamburger, M. Concepts and technologies for tracking bioactive compounds in natural product extracts: generation of libraries, and hyphenation of analytical processes with bioassays. *Nat. Prod. Rep.* **30,** 546–564 (2013).

41 Michel, T., Halabalaki, M. & Skaltsounis, A.-L. New concepts, experimental approaches, and dereplication strategies for the discovery of novel phytoestrogens from natural sources. *Planta Med.* **79,** 514–532 (2013).

42 Lang, G. *et al.* Bioactivity profiling using HPLC/microtiter-plate analysis: application to a New Zealand marine alga-derived fungus, Gliocladium sp. *J. Nat. Prod.* **69,** 621–624 (2006).

43 Johnson, T. A. *et al.* Natural product libraries to accelerate the high-throughput discovery of therapeutic leads. *J. Nat. Prod.* **74,** 2545–2555 (2011).

44 Johansen, K. T., Wubshet, S. G. & Nyberg, N. T. HPLC-NMR revisited: using time-slice high-performance liquid chromatography-solid-phase extraction-nuclear magnetic resonance with database-assisted dereplication. *Anal. Chem.* **85,** 3183–3189 (2013).

45 Wagenaar, M. M. Pre-fractionated microbial samples – the second generation natural products library at Wyeth. *Molecules* **13,** 1406–1426 (2008).

46 Strege, M. A. High-performance liquid chromatographic-electrospray ionization mass spectrometric analyses for the integration of natural products with modern high-throughput screening. *J. Chromatogr. B Biomed. Sci. Appl.* **725,** 67–78 (1999).

47 Bohni, N. *et al.* Integration of microfractionation, qNMR and zebrafish screening for the *in vivo* bioassay-guided isolation and quantitative bioactivity analysis of natural products. *PLoS One* **8,** e64006 (2013).

48 Nielsen, K. Fungal metabolite screening: database of 474 mycotoxins and fungal metabolites for dereplication by standardised liquid chromatography–UV–mass spectrometry methodology. *J. Chromatogr. A* **1002,** 111–136 (2003).

49 Nielsen, K. F., Månsson, M., Rank, C., Frisvad, J. C. & Larsen, T. O. Dereplication of microbial natural products by LC-DAD-TOFMS. *J. Nat. Prod.* **74,** 2338–2348 (2011).

50 Boswell, P. G., Schellenberg, J. R., Carr, P. W., Cohen, J. D. & Hegeman, A. D. A study on retention 'projection' as a supplementary means for compound identification by liquid chromatography-mass spectrometry capable of predicting retention with different gradients, flow rates, and instruments. *J. Chromatogr. A* **1218,** 6732–6741 (2011).

51 Ridder, L. *et al.* Automatic chemical structure annotation of an LC–MS(n) based metabolic profile from green tea. *Anal. Chem.* **85,** 6033–6040 (2013).

52 Konishi, Y. *et al.* Molecular formula analysis by an MS/MS/MS technique to expedite dereplication of natural products. *Anal. Chem.* **79,** 1187–1197 (2007).

53 Horai, H. *et al.* MassBank: a public repository for sharing mass spectral data for life sciences. *J. Mass Spectrom.* **45,** 703–714 (2010).

54 López-Pérez, J. L., Therón, R., del Olmo, E. & Díaz, D. NAPROC-13: a database for the dereplication of natural product mixtures in bioassay-guided protocols. *Bioinformatics* **23,** 3256–3257 (2007).

55 Holt, T. G. The design and development of an integrated natural products screening database. *J. Biomol. Screen.* **5,** 421–433 (2000).

56 Månsson, M. *et al.* Explorative solid-phase extraction (E-SPE) for accelerated microbial natural product discovery, dereplication, and purification. *J. Nat. Prod.* **73,** 1126–1132 (2010).

57 Singh, S. B. *et al.* Occurrence, distribution, dereplication and efficient discovery of thiazolyl peptides by sensitive-resistant pair screening. *J. Antibiot. (Tokyo)* **66,** 599–607 (2013).

58 Kim, H. J., Jee, E. H., Ahn, K. S., Choi, H. S. & Jang, Y. P. Identification of marker compounds in herbal drugs on TLC with DART-MS. *Arch. Pharm. Res.* **33,** 1355–1359 (2010).

59 Lane, S. *et al.* Toward single-calibrant quantification in HPLC. A comparison of three detection strategies: evaporative light scattering, chemiluminescent nitrogen, and proton NMR. *Anal. Chem.* **77,** 4354–4365 (2005).

60 Cremin, P. A & Zeng, L. High-throughput analysis of natural product compound libraries by parallel LC-MS evaporative light scattering detection. *Anal. Chem.* **74,** 5492–5500 (2002).

61 Hutchinson, J. P. *et al.* Comparison of the response of four aerosol detectors used with ultra high pressure liquid chromatography. *J. Chromatogr. A* **1218,** 1646–1655 (2011).

62 Adnani, N., Michel, C. R. & Bugni, T. S. Universal quantification of structurally diverse natural products using an evaporative light scattering detector. *J. Nat. Prod.* **75,** 802–806 (2012).

63 Vervoort, N., Daemen, D. & Török, G. Performance evaluation of evaporative light scattering detection and charged aerosol detection in reversed phase liquid chromatography. *J. Chromatogr. A* **1189,** 92–100 (2008).

64 Hasada, K. *et al.* Quantitative determination of atractylon in Atractylodis Rhizoma and Atractylodis Lanceae Rhizoma by 1H-NMR spectroscopy. *J. Nat. Med.* **64,** 161–166 (2010).

65 Pauli, G. F., Jaki, B. U. & Lankin, D. C. A routine experimental protocol for qHNMR illustrated with Taxol. *J. Nat. Prod.* **70,** 589–595 (2007).

66 Farrant, R. D. *et al.* NMR quantification using an artificial signal. *Magn. Reson. Chem.* **48,** 753–762 (2010).

67 Lang, G. *et al.* Evolving trends in the dereplication of natural product extracts: new methodology for rapid, small-scale investigation of natural product extracts. *J. Nat. Prod.* **71,** 1595–1599 (2008).

68 Wolfender, J., Queiroz, E. F. & Hostettmann, K. The importance of hyphenated techniques in the discovery of new lead compounds from nature. *Expert Opin. Drug Discov.* **1,** 237–260 (2006).

69 Jaroszewski, J. W. Hyphenated NMR methods in natural products research, part 1: direct hyphenation. *Planta Med.* **71,** 691–700 (2005).

70 Jaroszewski, J. W. Hyphenated NMR methods in natural products research, Part 2: HPLC-SPE-NMR and other new trends in NMR hyphenation. *Planta Med.* **71,** 795–802 (2005).

71 Williamson, R. T. *et al.* New diffusion-edited NMR experiments to expedite the dereplication of known compounds from natural product mixtures. *Org. Lett.* **2,** 289–292 (2000).

72 Cassani, J., Nilsson, M. & Morris, G. A. Flavonoid mixture analysis by matrix-assisted diffusion-ordered spectroscopy. *J. Nat. Prod.* **75,** 131–134 (2012).

73 Inokuma, Y. *et al.* X-ray analysis on the nanogram to microgram scale using porous complexes. *Nature* **495,** 461–466 (2013).

74 Genilloud, O. *et al.* Current approaches to exploit actinomycetes as a source of novel natural products. *J. Ind. Microbiol. Biotechnol.* **38,** 375–389 (2011).