Combined LC/UV/MS and NMR Strategies for the Dereplication of Marine Natural Products

Authors **Ignacio Pérez-Victoria, Jesús Martín, Fernando Reyes**

Affiliation Fundación MEDINA, Centro de Excelencia en Investigación de Medicamentos Innovadores en Andalucía, Granada, Spain

Key words
- marine natural products
- dereplication
- databases
- microfractionation
- LC/UV/MS
- NMR
- MS Gold
- MEDINA-HRMS

Abstract

Drug discovery from marine natural products has experienced a revival since the beginning of this century. To be successful in this field, rapid dereplication (identification of already known bioactive compounds) is essential in order to assess the chemical novelty of crude extracts and their fractions. Access to the appropriate state-of-the-art analytical instrumentation and to suitable databases is a fundamental requirement in such a task. A brief survey of the most robust LC/UV/MS- and NMR-based approaches employed for marine natural product dereplication is presented alongside a description of the procedures followed to achieve this goal in our research group.

Abbreviations

^{1}H-SF:	proton NMR structural features
MNP:	marine natural product
NP:	natural product
RT:	retention time
SPE:	solid phase extraction

Introduction

The discovery of promising drug leads and drug candidates from the sea has been the main driving force to keep exploring marine resources for bioactive compounds [1]. As a consequence, MNP chemistry has enjoyed a renaissance in the last 15 years [2]. The development of efficient dereplication strategies, thanks to the improvements in analytical technologies and to the availability of appropriate databases, has been critical in such a successful revival [3]. Dereplication means recognizing and eliminating the active substances already known in the early stage of the screening process [4,5]. In other words, it is the process of analyzing the active samples (crude extracts, fractions, etc.) identified in the preliminary biological screening using a combination of analytical separation and spectroscopic methods in order to identify known bioactive metabolites and eliminating unnecessary isolation work on already well-studied NPs [6]. It could be added that ideally the identified bioactive compounds should also explain the observed bioactivity in which the screening is based on (i.e., an antibacterial compound may not be an antifungal), otherwise other compounds in the sample should not be discarded as potentially interesting. Caution should be taken, within a drug discovery context, to not overlook such a fundamental relationship. Many would agree with some fellow MNP chemists who refer to *dereplication* as "an ugly word describing an ugly process" [7]. Such a statement just reflects the difficulties of this task due to the huge number of known MNPs (over 30000 entries in the current edition of the Dictionary of Marine Natural Products [8] and ~27500 compounds in the current release of MarinLit [9]) and to the chemical complexity of the samples (extracts and fractions), which frequently contain a large number of components covering a broad dynamic range of concentrations. Fortunately, during the last decade, the remarkable technological advances in analytical instrumentation, alongside the development of suitable databases, have enabled the rapid dereplication processes that are required to efficiently move forward in any drug discovery program based on MNPs. Dereplication approaches employed in other areas of NP chemistry (plants, terrestrial microbes, etc.) can easily be adapted to MNPs provided the proper caution on sample preparation before analysis is taken and the proper databases are employed [7,10–19]. Herein we present a survey of the most

received October 23, 2015
revised Dec. 22, 2015
accepted January 5, 2016

Bibliography
DOI http://dx.doi.org/10.1055/s-0042-101763
Published online March 22, 2016
Planta Med 2016; 82: 857–871

ISSN 0032-0943

Correspondence
Dr. Ignacio Pérez-Victoria
Fundación MEDINA
Avda del Conocimiento 34
18016, Armilla, Granada
Spain
Phone: + 34 958 99 39 65
Fax: + 34 958 84 67 10
ignacio.perez-victoria@medinaandalucia.es

Table 1 Relevant databases for MNP dereplication.

Database	Number of compounds						NMR data		
	Total	NP	Taxo.	Bioactiv.	UV λ_{max}	^{1}H-SF	δ	Spectra	HSQC/DEPT
CAS REGISTRY	66×10^6	~ 260 000	+	+	–	–	–	–	–
ChemSpider[a]	35×10^6	~ 70 000	–	–	–	–	–[b]	–	–
PubChem[a]	61×10^6	?	–	+	–	–	–[b]	–	–
REAXYS	2.7×10^6	> 215 000	+	+	–	–	–	–	–
UNPD[a]		229 358	–	+	–	–	–	–	–
DNP	> 272 000[c]	> 170 000[d]	+	+	+	+[e]	–	–	–
DMNP	> 30 000[c]		+	+	+	–	–	–	–
AntiMarin		~ 60 000	+	+	+	+	+[f]	–	–
AntiBase		42 950	+	+	+	–	+[g]	–	–
MarinLit		~ 27 500	+	+	+	+	+[h]	–	+[h]
StreptomeDB[a]		4041	+	+			+[h]		
SpecInfo	~ 360 000	?	–	–	–	–	+	+	–
ACD/NMR DB	322 000	~ 50 000	–	–	–	–	+	–	+
MICRONMR	730 000	~ 400 000[i]	+	–	–	–	+	–	–
CH-NMR-NP[a]		30 500	+[j]	–	–	–	+	–	–

[a] Public domain database; [b] It will be presented in the text software tools that employ the calculated NMR shifts of the structures contained in these databases; [c] Number of entries; [d] Number of NPs not considering derivatives; [e] Prof. John Blunt developed a version containing the ^{1}H-SF of the molecules contained in the DNP available up to release 22.1; [f] Not all entries contain chemical shift data (either experimental or calculated); [g] Entries without experimental chemical shift data display calculated data as an alternative; [h] Just predicted chemical shift data; [i] Information provided the vendor; [j] Biological source information is displayed in the hit results but cannot be searched for

robust and popular LC/UV/MS- and NMR-based approaches employed for MNP dereplication alongside a description of the procedures adopted at Fundación MEDINA for this task.

Databases

Appropriate databases play a major role in the dereplication of MNPs [7, 20, 21]. The ideal database should contain as much information on a compound as possible and should be searchable by substructure, structure identity/similarity, and spectroscopic identity/similarity. Desirable query fields include the trivial name, the molecular weight (MW), the accurate mass (or the monoisotopic mass), the molecular formula (MF), the taxonomic identification of the producing organism, the biological activity of the compound, and the UV absorption maxima. Those containing actual NMR spectra or chemical shifts (either experimental or calculated) are preferred. As an alternative, the concept of ^{1}H-SF developed by Professors John Blunt and Murray Munro has shown to be very effective in discriminating between alternative candidate structures in the dereplication process [22–24]. This approach searches for structural features that can be easily determined from inspection of the ^{1}H NMR spectrum of a compound. A list of the most relevant databases for MNP dereplication is presented in **Table 1**.

The most comprehensive compilation of information on NP substances is the Chemical Abstracts Service CAS Registry database [25]. A very appropriate access pathway to information contained in this database is via SciFinder [26]. The SciFinder interface allows searches in various ways to establish the previous occurrence or novelty of a compound, or its similarity to other known compounds, but surprisingly does not provide a direct search for all substances with a particular mass (this result though can be achieved by first carrying out a substructure search for all compounds containing C and then refining the search based on mass). REAXYS contains an extensive repository of experimentally validated data with a direct link to the relevant bibliographic references arranged by information type (spectra, bioactivity data, etc.) [27]. These features alongside the > 215 000 NPs included make it a very powerful database even though a search by accurate mass is not possible. There is a large number of databases with free access to chemical structures from various sources [28, 29]. Among them, PubChem [30, 31] (hosted by the National Institutes of Health, NIH) and ChemSpider [32] (hosted by the Royal Society of Chemistry, RSC) are the most useful. Both allow searches by substructure and structure similarity, though only ChemSpider allows for monoisotopic mass searches. The Universal Natural Products Database (UNPD) from Peking University was designed to be a comprehensive resource of NPs for virtual screening [33, 34]. This database can be queried by CAS registry number, chemical name, MF, and MW, thus being of some utility for dereplication. The Chapman & Hall/CRC Press Dictionary of Natural Products (DNP) is considered the gold standard of NP databases containing relevant bibliographic references for all compounds [35, 36]. The database contains information on biological source and bioactivity. It can be searched by accurate mass (apart from MW and MF) and UV maxima. Recently, Prof. John Blunt developed a version containing the ^{1}H-SF of the molecules contained in the DNP, available up to release 22.1, dramatically enhancing the dereplication utility of this database [37]. The Dictionary of Marine Natural Products (DMNP) is just a subset of data from the DNP based on the biological source of the compounds [8]. As such, it is obviously relevant for MNP dereplication. AntiBase 2014 covers NPs from microorganisms and higher fungi [38]. It includes descriptive data (MF, MW, CAS registry number), physicochemical data (melting point, optical rotation), some spectroscopic data for many of the compounds (UV, ^{13}C and ^{1}H-NMR, IR, and mass spectra), biological data (pharmacological activity, toxicity), information on origin and isolation, and a summary of literature sources. It also uses predicted ^{13}C NMR spectra (via SpecInfo [39]) for those compounds where no measured spectra are available. This database is becoming increasingly important for MNP dereplication since the overlap between "marine" microorganisms and "terrestrial" microorganisms can be difficult to determine. MarinLit is a database dedicated to MNP research. Originally established in the 1970 s by Professors John

Blunt and Murray Munro, it is currently hosted by the RSC [9]. It contains ~ 27 500 compounds. All records contain the usual bibliographic information and the database can be searched by querying substructure, ^{1}H-SF, calculated ^{13}C and ^{1}H NMR shift data, exact mass, chemical formula, or UV maxima. AntiMarin was available until very recently to subscribers of both AntiBase and MarinLit [40]. It contained data for compounds of both databases including ^{1}H-SF among the searchable fields and was probably the preferred database for dereplicating MNP among colleagues working in this field. The SpectromeDB is an outstanding public resource compiling the largest curated database of NPs isolated from *Streptomyces* species [41, 42]. It has been developed by the Pharmaceutical Bioinformatics group of Prof. Günther at the University of Freiburg. In addition to names and molecular structures of the compounds, information about source organisms, references, biological role, activities, and biosynthesis routes is included. The database includes virtual MS fragmentation patterns calculated with CFM-ID software [43] and predicted ^{1}H and ^{13}C NMR shifts generated with the NMR predictor tool from the command-line platform cxcalc [44]. StreptomeDB can be searched through queries on the previous fields plus MS peaks and ^{1}H or ^{13}C shifts, it is thus of great interest for dereplication of MNPs produced by marine actinomycetes.

While the previously described databases contain some spectroscopic data, or at least reference to the source of experimental spectroscopic data for a compound, there are other databases dedicated to the cataloging and/or calculation of spectroscopic properties. Access to these data can be particularly helpful in the investigation of MNPs. SpecInfo is a spectroscopic database whose primary aim is to assist with spectral interpretation and structure elucidation [39]. These functions are supported by tools for searching the database for compounds with NMR and/or IR spectra, or fragments of spectra, matching the experimental data. Additionally, compounds can be searched for using structures or substructures or other structurally related information such as CAS numbers. NMR spectrum prediction for a proposed structure is also an integral part of SpecInfo. These capabilities are supported by a knowledge base of 359 000 ^{13}C NMR spectra and 130 000 ^{1}H NMR spectra. Advanced Chemistry Development, Inc. (ACD/Labs) HNMR and CNMR Predictors permit the calculation of ^{1}H and ^{13}C NMR spectra from user-inputted structures [45]. These Predictors utilize algorithms based on more than 1.9 million assigned ^{1}H chemical shifts from more than 228 000 chemical structures and 2.7 million assigned ^{13}C chemical shifts from about 212 000 chemical structures. Use of these Predictors can be very helpful as a structural verification tool to assess the feasibility of a proposed structure. Of particular relevance to MNP dereplication are the internal databases in the Predictor packages. These contain the published chemical shift data for ~ 250 000 compounds. Presently, over half of the MNPs have their data included in the internal databases, and these data are being added to on a regular basis so that the proportion of MNPs contained in the internal databases will eventually be much higher. Currently, the ACD/Labs-calculated ^{1}H and ^{13}C NMR chemical shift data for all MNPs are accessible from within MarinLit, as described earlier. Another related product from this company, Structure Elucidator Suite, contains an internal library of over 2 million structural fragments from ~ 410 000 compounds [46]. The MICRONMR database was recently produced by Shanghai Micronmr Infor Technology Co., Ltd. (supported by Fudan University, Shanghai, P. R. China) [47]. In this database, ^{13}C NMR data of 730 000 organic compounds and related information (including bibliographic references) are indexed. About 400 000 entries correspond to NPs according to the vendor, though we think probably the synthetic derivatives of the NPs must be included in such a large estimation. The database can be searched by ^{13}C NMR data, compound name, genus, and species. It is thus a very powerful database for MNP dereplication when ^{13}C NMR data is available [48]. The CH-NMR-NP is a ^{13}C and ^{1}H NMR database for NP that was compiled from leading journals published from 2000 to spring 2014 [49, 50]. Important criterion to adopt the data was that ^{13}C shifts were given to all carbons in the chemical structure together with the description of ^{1}H shifts. The total number of the compounds is about 30 500 including 926 compounds related to NPs from the Spectral Database for Organic Compounds (SDBS) NMR database. The database was built by Dr. Kikuko Hayamizu [51] and it has recently been turned into a free of charge service released on the JEOL RESONANCE, Inc. website [52]. The compound name, atoms, MF, MW, ^{13}C and ^{1}H chemical shifts, ^{13}C no signal region, and structure (including partial structure) can be used as query items. The hit results include the biological source of the compound and the bibliographic reference from where the data were obtained. Other open-access NMR spectral databases such as NAPROC-13 [53, 54], MetIDB [55, 56], Spektraris [57, 58], and NMRShiftDB [59, 60] do not seem too useful for MNP dereplication, though they may find utility in other areas such as phytochemistry. Similarly, publicly available metabolomic databases, which include NMR data such as BMRB (Biological Magentic Resonance Databank) [61, 62], HMDB (Human Metabolome Database) [63, 64], MMCD (Madison Metabolomics Consortium Database) [65, 66], TOCCATA [67, 68], and related databases [69–72], are of no utility since they contain a negligible number of secondary metabolites and probably no MNPs at all [73].

Regarding databases containing mass spectral data, those with MS/MS information seem most useful. However, just a few libraries are currently available. The major reasons are the lack of requirements to publish MS/HRMS spectra and the lack of standardization of fragmentation energies between instrument manufacturers [17]. In any case, the two most popular ones are Massbank [74, 75] and METLIN [76, 77], covering ~ 10 000 compounds with their spectra. Although only containing relatively few NPs, the two libraries can be helpful to identify lipids, medium polar primary metabolites, vitamins, etc. Much more useful, however, is the in-house DTU (Technical University of Denmark) mycotoxin-fungal secondary metabolite MS/HRMS library [78], whose public part covers 277 compounds. It thus may be relevant for dereplicating samples derived from marine fungi. Also, the MS/MS database included in the recently developed Global Natural Products Social Molecular Networking resource (GNPS) created by the group of Professor Peter Dorrestein at the University of California, San Diego (UCSD) appears to be very promising [79]. With an emphasis on NPs of all biological origins, the database is community contributed and curated. This initiative aims to provide the definite public collection of MS/MS spectra of NPs. Currently, the database contains a significant number of entries contributed by leading research groups in MNPs at UCSD and the Scripps Institution of Oceanography, making it very relevant for MNP dereplication.

Finally, we would like to stress that dereplication is best carried out knowing the taxonomy of the producing organism [7], though this is not an absolute requirement [20, 21]. Within the databases included in **Table 1**, the coverage of taxonomy is good. Care should be taken though regarding the accuracy of the stated microbiological sources. Unfortunately, the primary bib-

liographic references may be mistaken due to the lack of proper and rigorous nucleic acid sequence data for unambiguous microbiological identification. To avoid unsuccessful or misleading dereplication due to erroneous taxonomy in the queried database, we suggest comparing the results obtained when approaching the dereplication problem at different levels of taxonomic ranks such as species, genus, and kingdom, apart from a parallel dereplication not accounting for any biological source information.

Analytical Separation and Microfractionation

Due to their complex chemical nature, the analysis of marine samples (crude extracts or rough fractions) mostly relies on liquid chromatography techniques (HPLC or UHPLC). The best choice for separation is reversed-phase (RP) chromatography, which suits the polarity of most drug-like secondary metabolites [16]. A low pH in the mobile phase is preferable, though care should be taken with the actual selection of the buffer/acidifier and its concentration to avoid chromatographic resolution and peak shape issues, ionization problems compromising the MS signal, and, finally, issues of chemical stability [17]. For the analysis of small and very polar compounds, techniques such as hydrophilic interaction chromatography (HILIC) may provide a satisfactory solution [80]. The RT (or retention index) of any compound is an intrinsic chromatographic property that can be employed as a query field for dereplication using in-house databases [81–83]. For compounds not included in such databases, the prediction of RT (under the same chromatographic conditions) is claimed to be a powerful tool [84], allowing, for example, to discriminate among dereplication candidates obtained from low-resolution mass spectrometry (LRMS) [85,86] or high-resolution mass spectrometry (HRMS) analysis [87–89]. The hyphenation of LC with detection based on UV/vis spectroscopy, MS, or even NMR provides a very convenient access to spectral information on the sample components [90,91]. On the other hand, both analytical and semipreparative HPLC can, at the same time, be employed to microfractionate crude natural extracts for direct assessment of bioactivity and active-peak identification [14,92,93]. Such a parallel biological/chemical profiling approach has been adopted for the preparation of MNP libraries suitable for high-throughput screening (HTS) [22,23,94–96] or even could be amenable for *in vivo* zebrafish-based assays [97–99].

Dereplication Based on Ultraviolet/visible Spectroscopy

UV/vis spectra are readily acquired using a diode array detector (DAD) as part of the LC or LC/MS examination of a crude extract or fraction. The spectra profiles, including the maxima (λ_{max}), provide information on the compound's chromophores, which can be used for database searching in different ways. Among the databases listed in the previous section, a number contain searchable UV maxima such as the DMNP or MarinLit. On the other hand, the data on λ_{max} for the compounds (mainly fungal metabolites and mycotoxins) contained in some in-house databases have been published and eventually may be useful when dereplicating marine fungi samples [82,100,101]. However, the whole UV spectrum contains more information than just the λ_{max} data and thus matching spectra is a superior and more definitive approach of dereplication. Unfortunately, no UV spectral databases that also contain other information essential to the dereplication process are available and the approach is just applied to in-house UV spectral databases developed within different research groups [7,82,102,103]. Spectral matching is carried out via algorithms, such as the X-hitting, which is designed for automated comparison of full UV spectra from LC-DAD analysis against a UV library of standards as well as spectra across samples [103–105]. Deconvolution of pure component spectra from overlapping LC peaks using multivariate curve resolution may significantly enhance the approach [106]. This allows for both the identification of known compounds as well as new compounds with UV spectra similar to known compounds [107]. Currently, DAD data is used as second (or complementary) criteria after an MS-based search. In favorable cases, UV/vis spectral chromophores can provide a means to differentiate compounds with the same elemental composition and can be highly valuable in dereplication for exclusion or confirmation of candidates during a database search.

Dereplication Based on Mass Spectrometry and Molecular Formula

LRMS was initially employed for LC-MS-based dereplication against local databases [82,102,103] or low-discriminating dereplication based on MW searches on most of the databases previously listed. The advent of high-resolution instruments (HRMS) has significantly improved dereplication approaches, allowing for the use of generic databases through MF searches or the use of accurate mass searches in specific databases such as DMNP and ChemSpider. Under electrospray ionization conditions (the most common one), care must be taken though to properly assign the right molecular ions [100] and also to correctly determine the MF, even when mass accuracies < 1 ppm are obtained [108]. In fact, we favor databases allowing experimental accurate mass searches since they may indirectly assist in MF determination. The first LC-LRMS dereplication methods using local databases have now evolved to employ LC-MS/MS [109], HPLC-DAD-HRTOFMS [100], or UPLC-DAD-HRMS-MS/MS [101]. The recently developed "aggressive dereplication" approach employing UP-LC-DAD-QTOF is based on accurate mass, isotopic patterns, and preferably selective adducts used for large batch searches of possible metabolites (up to 3000 compounds), e.g., based on all compounds described by a single genus. Yet, it returns false positives that need to be sorted away. The approach is currently not suited for organisms with limited taxonomic information [110]. A more robust strategy from the same authors uses the same equipment to integrate tandem MS/HRMS data and an in-house library containing 1300 compounds (mostly fungal) for unambiguous dereplication [111]. Interestingly, for compounds not contained in the library, the aggressive dereplication approach may still be applied. All these LC/MS-based approaches integrate DAD, as mentioned in the previous section, to take advantage of the cheap and sometimes unique dereplication power of UV/vis spectroscopy. On the other hand, the use of chromatography allows querying the RT in the local databases. Interestingly, Boswell et al. described that since MS data and RT are orthogonal, compounds can be identified more efficiently from a combination of RT and LRMS rather than HRMS alone [85]. A completely different approach is based on the computational prediction of MS/MS spectra or fragmentation pathways. The limited coverage of available experimental databases has led to an interest in computational

methods for predicting reference MS/MS spectra from chemical structures [112–114]. Commercial packages, such as Mass Frontier (ThermoScientific) [115] and ACD/MS Fragmenter (ACD/Labs) [116], are rule-based and employ thousands of manually curated rules to predict fragmentations using non-published proprietary algorithms. Alternatively, there are open-access resources providing *in silico* fragmentation tools that may be extraordinary useful for dereplication purposes: MetFrag software uses high-resolution tandem mass spectra of metabolites as a first and critical step for the identification of a molecule's structure [117]. Candidate molecules of different databases (including PubChem and ChemSpider) are fragmented *in silico* using a combinatorial fragmenter algorithm based on the bond disconnection approach and matched against the mass to charge values very rapidly. A score calculated using the fragment peak matches gives hints to the quality of the candidate spectrum assignment [118]. An alternative method, FingerID [119], takes advantage of the increasing number of available MS/MS spectra by applying machine learning methods to this task [120]. This program uses support vector machines to predict a chemical fingerprint directly from an MS/MS spectrum, and then searches (e.g., in PubChem) for the metabolite that most closely matches that predicted fingerprint. Competitive fragmentation modeling (CFM) is a recently introduced probabilistic generative model for the MS/MS fragmentation process that uses machine-learning techniques to learn its parameters from real data [43]. It can also be applied for the putative metabolite identification task (CFM-ID) ranking possible structures from PubChem for a target MS/MS spectrum [121, 122]. The StreptomeDB contains virtual fragmentation information obtained via CFM-ID [42]. The MAGMa method [123] has also been recently developed to automatically process and annotate the LC-MS^n data sets (including MS/MS data) on the basis of candidate molecules from chemical databases, such as PubChem [124]. Multistage MS^n spectral data is automatically annotated with hierarchical trees of *in silico* generated substructures of candidate molecules to explain the observed fragment ions [125], and alternative candidates are ranked on the basis of the calculated matching score [126]. Even more recently, the CSI (Compound Structure Identification):FingerID method for searching a molecular structure database such as PubChem using MS/MS data has been released [127]. This method combines computation and comparison of fragmentation trees with machine-learning techniques for the prediction of molecular properties of the unknown compound [128], and shows significantly increased identification rates compared with all existing state-of-the-art methods for the problem. All these public resources (MetFrag, FingerID, CFM-ID, MAGMa, CSI : FingerID plus others not listed here) have a great potential for dereplication of MNPs just limited by the actual number of these type of compounds contained in the queried databases (PubChem, etc.). We are currently evaluating their dereplication utility in real life situations related to projects dealing with MNPs and microbial and plant secondary metabolites, and the results will be published elsewhere.

Finally, the elegant and sophisticated new networking MS/MS approaches appear to be quite potent [129], however, they still seem too complex and time-consuming for widespread use among MNP chemists, at least in the short run.

Nuclear Magnetic Resonance-Based Approaches to Dereplication

▼

NMR spectroscopy provides much richer structural information on a compound than MS or UV. Rigorously speaking, LC-UV-MS (MS) does not necessarily confirm unequivocally the identity of known compounds because of the possibility of regioisomers or stereoisomers occurring. RTs from in-house databases may diminish this problem, but undoubtedly the definitive answer is given by NMR spectroscopy. For this reason it may be essential for successful dereplication. Additionally, NMR facilitates the identification of the structural class a putative new compound belongs to. The hyphenated techniques HPLC-NMR [130] and HPLC-SPE-NMR [131] enable structure determination of NPs directly from small amounts of extracts overcoming preparative-scale isolation and are currently typically integrated with parallel UV and MS detectors [132]. Current probe technology allows acquiring proton spectra (CapNMR™ probe) [133–136] or even 2D heteronuclear correlation experiments (MicroCryoprobe™) [137] with just micrograms of a sample obtained from analytical or semiprep HPLC fractionations [138]. Such an improvement in sensitivity has revolutionized the role of NMR in dereplication during the last decade [15]. Three different approaches are typically followed when using NMR for dereplication purposes: i) Searching spectra (or chemical shift lists) in libraries of experimental spectra (or databases including assigned chemical shits), ii) searching a library of virtual spectra (or databases including calculated chemical shifts), and iii) searching for structural features that can be easily determined from inspection of the proton spectrum of a compound in a database of compounds allowing queries by ^{1}H-SF.

i) A proton spectrum may be searched against an in-house library of spectra acquired under the same or different solvent/magnetic field. Software tools such as ACD/Spectrus DB [139], MNova DB [140], or Bruker AMIX [141] allow searching for peaks, multiplets, spin systems, regions, or even whole spectra similarity. It has been shown that ^{1}H NMR spectra, obtained through analytical HPLC separation of extracts with time-sliced SPE trapping of eluting compounds and acquired with a microcryoprobe, can be used for dereplication purposes by matching an in-house database using an algorithm developed and operating under Matlab [142]. On the other hand, the spectra contained in the commercial SpecInfo database by Wiley can be conveniently searched for in installations under ACD Labs software [143] or the KnowItAll NMR Spectral Library by Bio-Rad [144], which also includes the ^{13}C NMR spectra collection of Wolfgang Robien [145]. Regarding the use of chemical shift lists, in a recent work, the combination of a centrifugal partition extraction fractionation method with ^{13}C NMR and hierarchical clustering analysis (HCA) for pattern recognition of ^{13}C signals across spectra of the fraction series enabled the direct deconvolution of each of the main metabolites signals and their dereplication after searching a database with assigned experimental carbon shifts created in ACD Labs software [146]. A similar approach has also been reported using HPLC fractionation and the MICRONMR database [47], showing again how ^{13}C chemical shifts of mixture components can be determined without having to purify them thanks to HCA [48]. The internal databases of ACD Predictor packages can be searched by chemical shifts, shift correlations (H/C), or spin systems and are of particular relevance for dereplication due to the large number of MNPs included [45]. Proton and carbon shifts can also be used as query

in the CH-NMR-NP database [52], probably the best public resource for NMR-based dereplication of MNPs.

ii) In the absence of experimental spectra libraries or databases of assigned experimental chemical shifts, the use of structures databases to create libraries of virtual spectra (or databases with calculated shifts) appears as a promising and efficient alternative. Since the size of any in-house library will typically be just a small fraction of the known number of MNPs, such an approach is worth exploring. The Natural Products Chemistry Department at Merck Research Laboratories proposed the use of similarity searches over databases of estimated ^{13}C NMR spectra for NP dereplication and identification of the structural class of novel compounds [147]. The same lab later developed a database with calculated ^{13}C/^{1}H-^{13}C (correlation) spectral lists for 11 673 NPs, showing the promising efficiency of querying the peaks found on HSQC spectra (bearing in mind its short acquisition time compared with ^{13}C direct detection) [148]. The FindIt module of the NMRanalyst™ software contains (in the March 2008 release) over 14.5 million PubChem structures and the predicted proton and carbon shifts for these structures [149]. The database can be searched using ^{1}H, ^{13}C, or $^{13}C_H$ (protonated carbon) shifts plus MW or MF as options; interestingly the highest correct identification rate seems to be encountered with the ^{1}H, $^{13}C_H$, MF input combination [150]. The latest release of the ACD software Structure Elucidator Suite includes ca. 22 million structures from ChemSpider (those containing C, H, O, N, S, P, F, Cl, Br, and I) with their predicted NMR shifts [46]. The utility of this database to dereplicate NPs using minimal NMR data (proton, HSQC and COSY) combined with MF composition range and structural fragments has been recently shown [18]. The MarinLit database also contains calculated ^{13}C and ^{1}H NMR shifts and HSQC/DEPT provided by ACD/Labs [9]. The outstanding public domain resource CSEARCH for PubChem has been developed by Wolgang Robien to identify compounds using the predicted ^{13}C NMR spectra of 61 million PubChem structures (calculated with the CSEARCH Neural Network technology) and a spectral appearance in hierarchical order (SAHO) searching algorithm [151] including ranking of resulting hit list [152]. The SpectromeDB includes predicted ^{1}H and ^{13}C NMR shifts that can be searched for [42] and, as already mentioned, it is of great relevance for the dereplication of MNPs produced by marine actinomycetes.

iii) The last strategy uses databases with the capability of searching for the actual numbers of functional groups contained within a molecule. The first approach of Sidebottom and coworkers was based on searching a text file that links each structure with its MW and an exact count of the number of methyl, methylene, and methine groups it contains. Analysis of such a text file, constructed from a database containing more than 126 000 NP structures (combining DNP and Beilstein), revealed that these data, readily measured using MS and NMR spectroscopy (1D proton and HSQC), are highly discriminating [12]. More elaborate and useful was the approach of Professors John Blunt and Murray Munro to include in a structure database the number and type of methyl groups, alkenes, carbinol protons, acetal, formyl, acetyl, amide, imine, aromatic substitution patterns, sp^3 methines, sp^3 methylenes, and sp^2 protons as searchable fields [7, 20–24]. They based their design on the fact that certain structural features in a ^{1}H NMR spectrum are immediately obvious and do not need any interpretation to know what they are. A simple inspection of the spectrum and integrals immediately allows the identification of many of the classes of functional groups listed before without needing to consider any relative connectivity. Such development of a pattern recognition NMR database (also known as ^{1}H NMR structural features, ^{1}H-SF) does not rely on the assignment of chemical shifts or analysis of correlations, but it is impressively effective in discriminating between alternative candidate structures in the dereplication process [20]. This is because the probability of compounds having identical combinations of ^{1}H-SF is low, and if these data are also taken together with MW, MF, and UV data, unique search patterns are generated that can quickly establish even the putative novelty of an isolated compound. Currently, MarinLit and old versions of AntiMarin and the DNP NMR feature databases can be searched by ^{1}H-SF and, thus, are without any doubt the most potent databases for NMR-assisted dereplication of MNPs.

Fundación MEDINA Dereplication Workflow

Fundación MEDINA [153] works on various drug discovery programs in different therapeutic areas [154–160]. We aim to discover new lead compounds from our proprietary library of microbial (fungi, actinomycetes, and other bacteria) NP extracts. Typically these extracts are generated from microorganisms cultured in small volumes (10–50 mL). In any primary screening campaign, the hits found are submitted to an early LC-UV-LRMS dereplication process using an in-house database. Samples not discarded and that are suggestive of containing active compounds with a possible novelty (for not being present in the in-house library or not having been clearly identified as known components after further LC-HRMS analyses) are chosen and the corresponding microorganisms are fermented at a higher volume (typically 100 mL). After confirming the bioactivity of the newly generated extracts, these are further fractionated typically by reversed-phase semipreparative HPLC. Active fractions are then submitted to LC-DAD-HRMS analysis and dereplication against an in-house LC-HRMS library or using a combination of taxonomy, UV info, accurate mass, and MF searches against the DNP (or any other public access structure database). The dereplication candidates are further confirmed after NMR analysis of the fraction, which allows unambiguous annotation of both in-house LC-HRMS and LC-UV-LRMS libraries and populates the in-house NMR spectra database. Alternatively, we also employ HRMS/MS analysis combined with *in silico* fragmentation tools for confirming the dereplication candidates obtained after the previous LC-DAD-HRMS step. All compounds from the active samples not dereplicated in this workflow will undergo a bioassay-guided isolation and structure elucidation process. Afterwards, they are stored in the in-house libraries whether they turn out to be novel or not. This dereplication workflow is being used in the context of the PharmaSea project [161] – a collaborative European consortium looking for novel bioactive molecules from marine organisms, including deep-sea sponges, bacteria, and fungi, in the different screening campaigns with samples derived from our collection of marine microorganisms and from other partners in the consortium.

Early-stage dereplication by LC-UV-LRMS

Active extract analysis is undertaken with a standard 10-min reversed-phase gradient chromatographic run on an Agilent 1100 single quadrupole LC-DAD-MS system (MSD), collecting mass spectra in both the positive and negative modes [158]. Database searching is performed using an in-house developed application inherited from Merck & Co., the "MS Gold" dereplication soft-

ware, and the spectral libraries from its former Natural Products Chemistry department [102, 103]. In MS Gold, the DAD (UV-vis) spectra, RT, and positive and negative mass spectra of the samples are compared to the corresponding LC-UV-MS data of known microbial metabolites stored in the proprietary database. This is the Fundación MEDINA reference library which contains annotated secondary metabolite data obtained under identical conditions to those for the samples under analysis; the library includes 405 fungal metabolites and 478 metabolites from bacteria and actinomycetes, the original Merck library size has increased 25% with the compounds annotated at Fundación MEDINA since its origin in 2009. Such a library is dynamic and is continuously populated with NPs identified in our different drug discovery programs.

Data extraction procedures from the LC-DAD-MS analyses involve the use of the AMDIS (Automated Mass Spectral Deconvolution & Identification System) tool developed by the National Institute of Standards and Technology (NIST) for the extraction of pure component MS spectra from complex chromatograms [162, 163]. Peaks from the UV/vis LC trace at 210 nm are detected by integration in the acquisition instrument software (Agilent ChemStation), but no spectra deconvolution of overlapping LC peaks is carried out. For identification of components and data combination, the UV and MS RTs are standardized for the system offset and corrected based on an external standard. For each LC-DAD-MS injection, a list of components is created using three sources: all UV_{210} peaks as detected in the Agilent Chemstation and all MS peaks as reported by AMDIS for both positive and negative ion modes. The data are automatically captured for each component and if no UV_{210} peak is integrated, a UV spectrum is extracted for each deconvoluted AMDIS component at its precise RT. The data are combined for each component based on the set of data that appears within a small time window (**Fig. 1**). MS Gold is written with Visual Basic 6.0 and SQL Server 2000 [103].

For the spectral library setup (containing UV/DAD, positive and negative LRMS spectra), the components associated to known molecules (identified after further HRMS and NMR in our dereplication workflow as shown later) are stored in the database and flagged as members of the spectral library (**Fig. 1**). The library contains identification information and characteristic analytical data for all samples/components analyzed. When the compound is not fully characterized, a tentative status is assigned within the library. When the structure of the component is identified, the database record is marked as "fingerprint". The trivial name and registration information are attached. The extracted component data (RT, UV, and LRMS spectra) of every new sample analyzed by LC-DAD-LRMS are searched against such a library using a proprietary algorithm that includes a composite rating function to weigh the match ranking of each of the queried properties (RT and UV/MS spectra) [103]. As an example, **Fig. 2** shows the dereplication of ikarugamycin in an extract from the culture broth of a marine *Streptomyces zhaozhouensis* strain [164]. MS Gold also allows batch searching to process sets of samples, a very appropriate feature within an HTS context of NP drug discovery.

Advanced dereplication by LC-HRMS

Active extracts and their active fractions obtained by semiprep HPLC that have not been dereplicated with MS Gold are submitted to LC-DAD-HRMS analysis using the same chromatographic conditions and system as described in the previous stage, but acquiring the HRMS spectra with a Bruker maXis QTOF mass spectrometer (where MS/MS can also be acquired) [158]. By default, acquisition is performed in the ESI+ mode, switching polarity for cases where no ionization is achieved. Using a similar strategy to what MS Gold is based on, we have created a new in-house tool solution called "MEDINA-HRMS" written with Visual Basic.NET and ORACLE 11 g. Extraction of pure component HRMS spectra from the raw data is carried out by the instrument software (Bruker DataAnalysis) and the accurate mass for the extracted components is interpreted internally by the application based on the same well-established algorithm for molecular weight assignment employed by MS Gold [165]. Unfortunately, UV/vis obtained in the DAD detector cannot be automatically extracted to be incorporated in the database and thus the search algorithm is only based on matching both RT and the interpreted accurate mass of the pure component extracted by the software within a narrow tolerance window (**Figs. 3** and **4**). Nevertheless, when required, any UV/vis spectrum can be retrieved manually for its inspection to get information on the absorption maxima or even for eventual comparison with the spectra stored in MS Gold.

The MEDINA-HRMS library is automatically populated with the RT, accurate mass, and HRMS spectra of all the pure components extracted from all the samples that are analyzed. When the pure component is identified (after accurate mass/MF searches in databases or further NMR-based dereplication), its trivial name and actual MF are annotated and automatically propagated to all past and future samples containing the same specific component. Batch search for studying sets of samples is no longer needed in MEDINA-HRMS because of its design as a relational database.

For those pure components not already flagged with their name in the MEDINA-HRMS library, identification is carried out typically with the following strategy: The interpreted accurate mass (an experimental accurate mass) is searched for in the DNP (which contains the calculated accurate mass based on MF of the compounds). The DNP accurate mass value(s) within the tolerance range retrieves the first hit list of candidates and, at the same time, the likely MF of the compound. If the molecules in the hit list contain characteristic λ_{max}, the experimental UV (DAD) spectrum is checked for its compatibility with the reported absorption maxima. Additionally, the taxonomic (biological source) information listed in the DNP is also taken into account for hit list refinement.

Alternatively, we also search the interpreted accurate mass in ChemSpider or the determined MF in this same database or PubChem for further exploration of the hit lists obtained. The less common the MF is, the smaller the number of candidates retrieved and the easier the manual inspection of their structures. Such an approach is obviously easier the higher the MW of the unknown component. Additionally, a complementary and somewhat more robust strategy based on the acquisition of HRMS/MS spectra for the target unknown component is followed. This tandem MS spectra are obtained with the same LC-DAD-HRMS system described before, just selecting the target parent ion to be fragmented by collision-induced dissociation (CID). The experimental MS/MS spectrum obtained is first used for MF confirmation (or modification) using the vendor's application SmartFormula 3D [166]. The obtained peaks (fragments) list is then used as query input for searching ChemSpider and PubChem using the MetFrag Web tool [118] based on *in silico* fragmentation for the computer-assisted identification of metabolites [119]. The ranked list of candidates provided by MetFrag is inspected manually and compared with the hit list retrieved by DNP, and many times (provided the compound is included in ChemSpider or

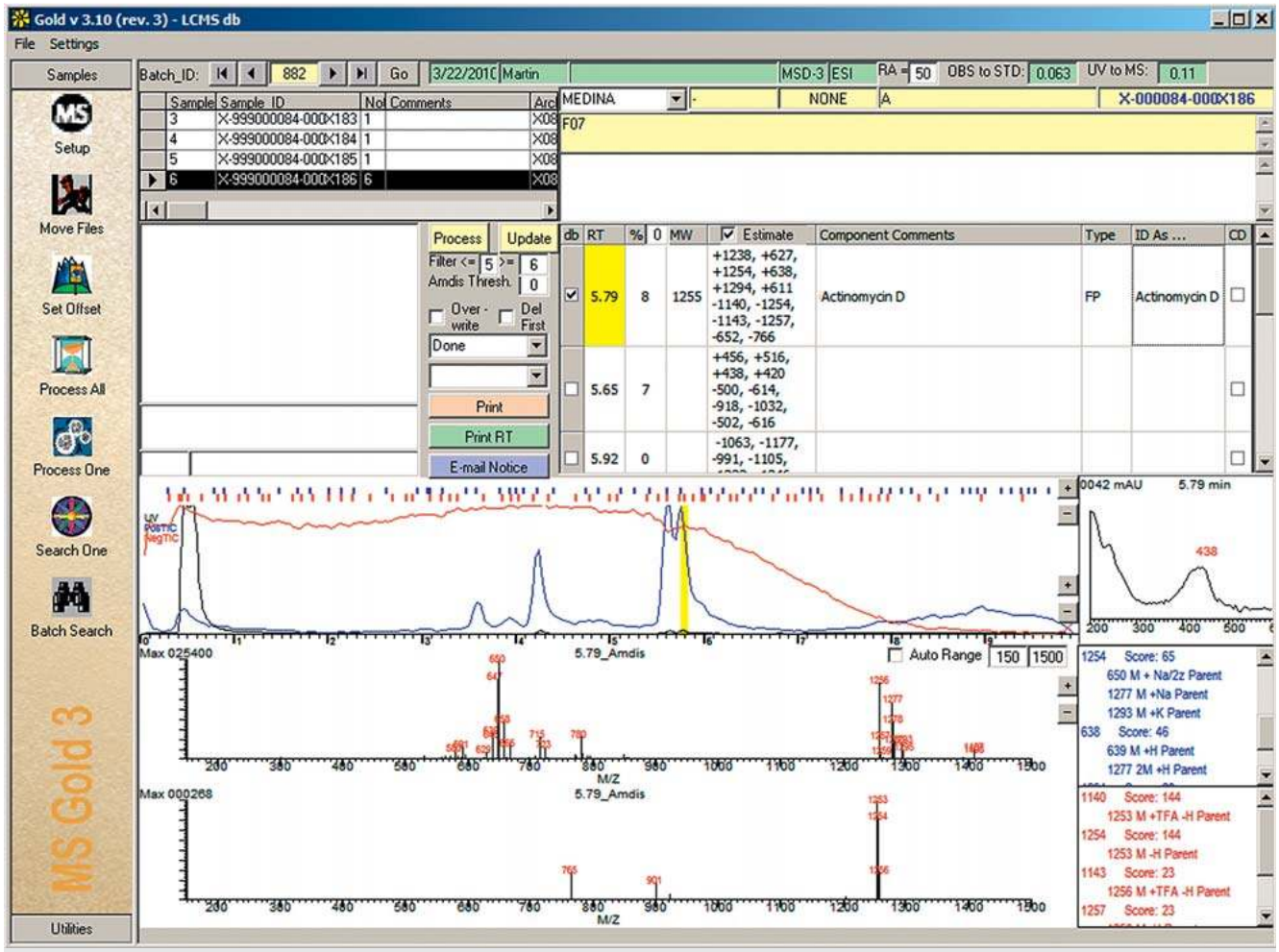

Fig. 1 Screenshot of a component (actinomycin D) selected as a fingerprint in the MS Gold library. (Color figure available online only.)

PubChem) the right answer is found using this strategy. Interestingly, using MetFrag and PubChem databases we were able to dereplicate an unknown component that turned out to be JBIR-34, a remarkable nonribosomal tetrapeptide possessing an unusual 4-methyloxazoline moiety and found in a marine sponge-associated actinomycete [167], the same year of its publication (2010) even though the structure was obviously not contained in the actual DNP release edition for that date. Such an example demonstrates an added value of these large public chemical compound databases that are updated with new entries much more frequently than the DNP (which offer two releases each year). Finally, it is worth mentioning that we likewise use dereplication by LC-HRMS as an assessment of putative novelty in cases where no match for the interpreted MF is found in DNP, ChemSpider, and PubChem databases [168].

Definite dereplication by NMR

Active fractions obtained by semiprep HPLC that have been already submitted to LC-DAD-HRMS-based dereplication are further analyzed by NMR for confirmation (or modification) of the dereplication candidate(s) and population of the in-house NMR spectra database. Likewise, bioactive fractions that could not be dereplicated (or had too many dereplication candidates) after LC-DAD-HRMS are dereplicated *de novo* based on NMR. The state-of-the-art NMR equipment available at Fundación MEDINA is comprised of a Bruker AVANCE III 500 MHz spectrometer equipped with a 1.7-mm TCI MicroCryoprobe™ enabling the acquisition of NMR spectra with a few micrograms of sample [139]. Typically, the dried bioactive semiprep HPLC fractions are reconstituted in deuterated solvent and submitted to routine ^{1}H and HSQC acquisitions. Identification of the target compound is carried out usually with the following strategy: The structural features easily identified in the proton and HSQC spectra are looked for in the dereplication candidate(s) obtained after LC-DAD-HRMS analysis. When agreement is observed, the spectra reported in the bibliography are compared with those of our sample for further confirmation. When no compatibility is observed between the acquired NMR spectra and the dereplication candidate(s), a *de novo* identification based on NMR is pursued. Such a dereplication process is likewise applied for those cases where no dereplication candidates (or too many hits) were obtained after LC-DAD-HRMS analysis, as indicated before. In a first stage, the structural features of the unknown (which are easily observed in the NMR spectra) are entered as query alongside the MF or MW in the DNP NMR features database [37]. The structure of the retrieved hits is inspected manually and the reported spectra for the compatible structures are again compared with those of our sample. Querying the ^{1}H-SF version of the DNP is extraordinarily effective to discriminate among possible candidates sharing the same MF. Interestingly, the same strategy, but using SF

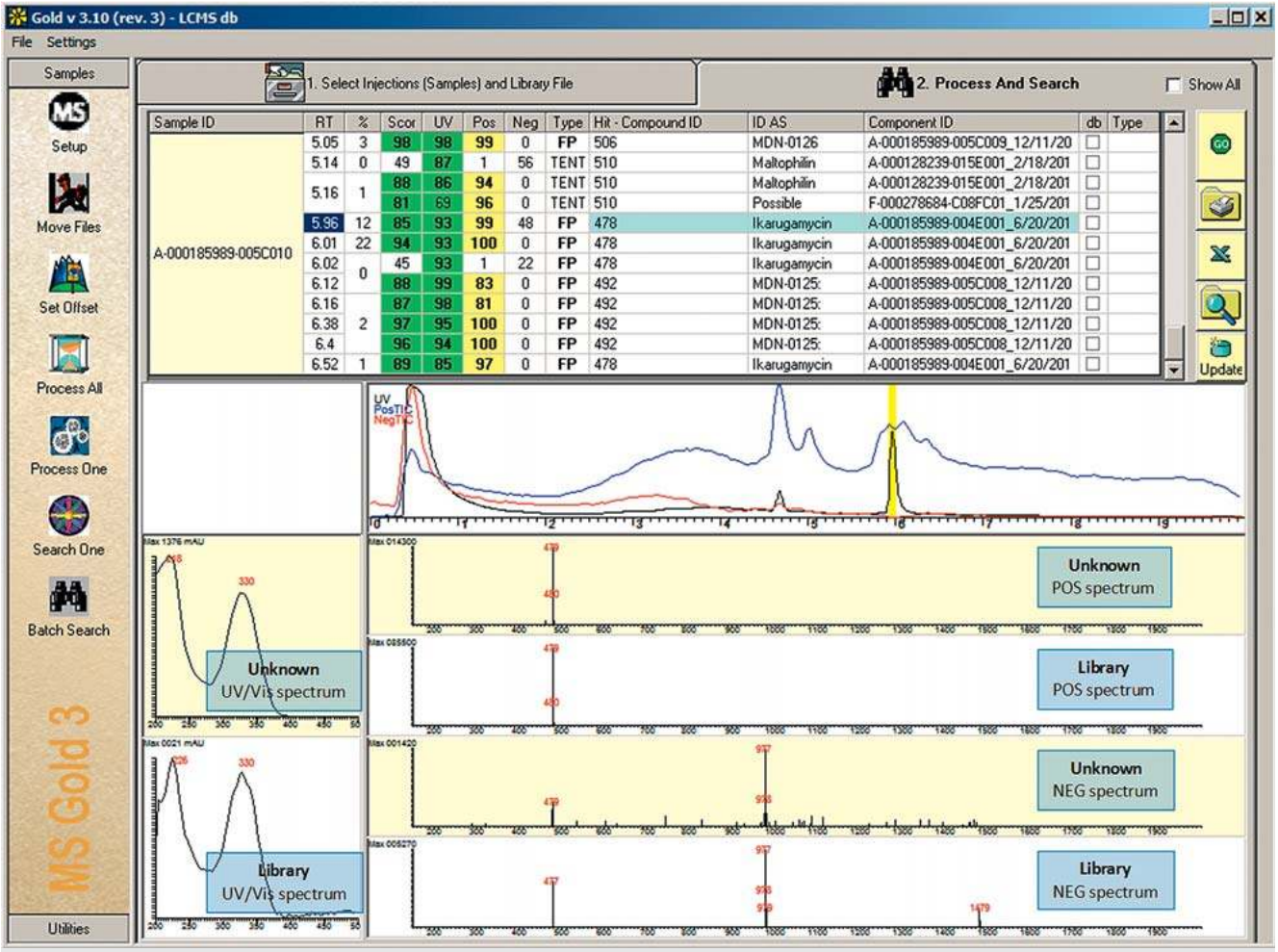

Fig. 2 Screenshot of MS Gold showing an example of ikarugamycin dereplication (component highlighted in yellow in the chromatogram) in an extract from the culture broth of a marine *Streptomyces zhaozhouensis* strain. (Color figure available online only.)

ranges and MF or MW ranges as complementary query fields in the same database, may allow, in a very easy way, the establishment of the structural class the unknown belongs to. In some cases we find in this way that the non-dereplicated unknown, after the LC-DAD-HRMS stage, turns out to be a simple derivative of a known NP, for example, carrying (or lacking) just an extra methoxyl or acetyl group. In a similar fashion, searching by similarity, the proton spectrum of the unknown against our in-house NMR spectral library (constructed in ACD/Spectrus DB software [139]) provides immediate information about possible structural relationships of the unknown and the NP contained in our library. As an alternative approach we also employ the FindIt module of the NMRanalyst™ software for searching over 14.5 million PubChem structures (and their predicted proton and carbon shifts) [149,150]. We usually employ ^{1}H, $^{13}C_{H}$, and MF as initial query fields, extending to MF or MW ranges in a second search round for also catching up compounds similar to the unknown. Recently, we have also started to explore the search of experimental chemical shifts of our unknowns for NMR-based dereplication using public resources such as CH-NMR-NP [52], StreptomeDB 2.0 [42] and CSEARCH for PubChem [152], the preliminary results and performance we have obtained so far are very satisfactory. Within the PharmaSea consortium [161], the cheminformatics company ACD/Labs has provided us with access to the internal databases in the ACD/C+H NMR Predictor packages [45]. Since these databases contain over half of the known MNPs, we are now starting to use them as a complementary approach for NMR-based dereplication in the context of this project. Both MS Gold and MEDINA-HRMS are annotated with all the compounds ultimately identified (or even elucidated) by NMR. In this sense, it is worth mentioning that unambiguous NMR-based dereplication may require, in some instances, the acquisition of a full panel of 2D NMR spectra. As an example of this, we describe here how we dereplicated the cyclic octapeptide surugamide A [169] in a sample from a marine-derived *Streptomyces* sp. in the context of the PharmaSea project. The extract obtained from the culture of this microorganism displayed moderate antibacterial activity. LC-DAD-LRMS analysis revealed the presence of one main component not included in our in-house library MS Gold. After LC-DAD-HRMS analysis of the bioactive fraction obtained by semiprep HPLC of the extract, the putative MF of the target compound was established and just one candidate with such a MF, named surugamide A [169], was obtained after searching the DNP database. Searching the MF in ChemSpider revealed the presence of another isomeric peptide, named champacyclin [170], with differences in the amino acid sequence and chirality compared to surugamide A. We then ran an HRMS/MS analysis and could check that only the sequence of surugamide A was

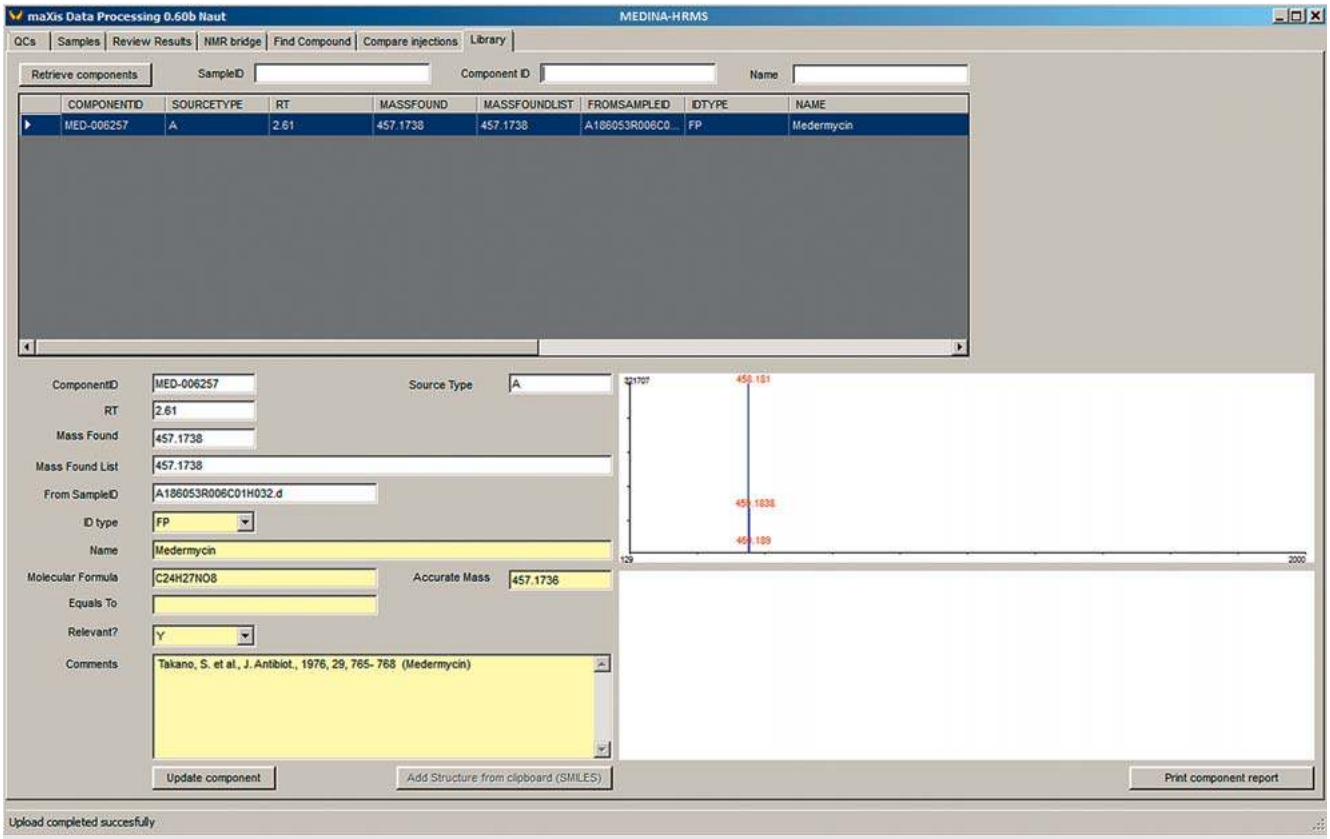

Fig. 3 Screenshot of a library ID component (medermycin) in MEDINA-HRMS. (Color figure available online only.)

compatible with the observed fragments. Since leucine and isoleucine are isobaric, and due to the fact that an unambiguous establishment of the cyclic peptide sequence sense (C to N or N to C) is impossible by MS-based approaches and that we might even be dealing with a new stereoisomer of surugamide A, a full panel of 2D NMR experiments (including COSY, HSQC, TOCSY, NOESY, and HMBC) was run to unambiguously determine the constituent amino acids and their actual arrangement in the peptide based on the sequential NOEs observed and the key interresidue HMBC correlations. After such an analysis, the connectivity of our unknown molecule was confirmed to match that of surugamide A. The final identity of the amino acid chirality was established after detailed comparison of the proton and carbon chemical shifts displayed for our unknown and those reported for surugamide A [169]. The extraordinary sensitivity of our MicroCryoprobe allowed the acquisition of the set of 2D spectra with submilligram amounts of sample (the bioactive fraction reconstituted in deuterated solvent).

In general, the selected bioactive semiprep HPLC fractions are dominated by a main component (ca. 75–90%) and identification of the NMR signals of the target compound is straightforward. Nevertheless, we have also faced scenarios where such fractions are comprised of a few components present in a similar ratio with respect to each other. These cases require further work for deconvoluting the NMR signals of each mixture component and we follow three different approaches: i) The first is based on further purification of the original semiprep HPLC fraction using different chromatographic conditions (changing gradient, mobile phase solvents, or even using a column with a different stationary phase). Ideally, the mixture components of the original fraction will have now been resolved and since the newly generated fractions are submitted again for bioassay, the correlation compound-activity can likewise be established unambiguously. The NMR spectra of the new bioactive sample are this way much simpler than the spectra of the original fraction containing now essentially one component. ii) An additional fractionation run by semiprep HPLC is not always feasible due to the small amount of material or to difficulties for achieving chromatographic resolution. We resolve these issues by analytical HPLC-DAD-SPE-NMR. With this technique, we work in the so-called tube transfer NMR mode (ttNMR) [171]. In this mode, the analytes trapped on the SPE cartridges are eluted with deuterated solvent into the capillary NMR tubes (HPLC-SPE-tube transfer NMR, HPLC-SPE-ttNMR) rather than being eluted into the flow cell of an NMR flow probe. That way an aliquot of the tube content can be analyzed by LC-DAD-HRMS and the remaining material can be easily resubmitted to the bioassay after NMR analysis. iii) The last approach is based on using pulsed field gradient (PFG) diffusion NMR spectroscopy for virtual separation of the sample components according to their molecular size [172, 173]. We have observed that this approach is feasible for mixtures containing just a few components (such as semiprep HPLC fractions) and that it is easier the higher the differences in MW among them. We have designed a research proposal to explode this approach of dereplication [174]. So far, we have observed that it is frequently possible to correlate exact molecular masses (or MFs) of the different mixture components (obtained in the previous LC-DAD-HRMS analysis stage) with the estimated molecular weights from diffusion experiments [175]. That way we can correlate a MF with a diffusion deconvoluted NMR spectrum of a mixture component and

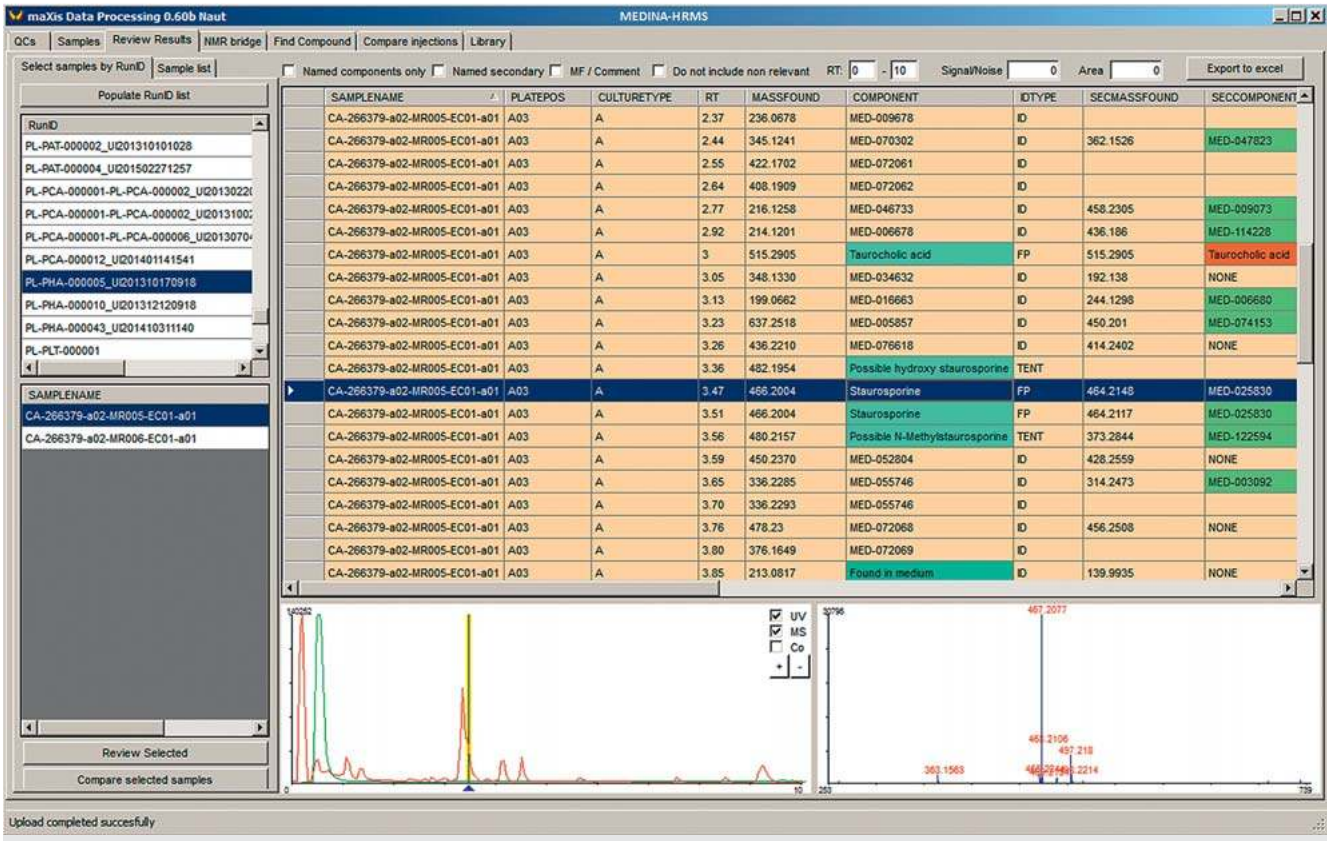

Fig. 4 Screenshot of MEDINA-HRMS showing an example of staurosporine dereplication (component highlighted in yellow in the chromatogram) in an extract from the culture broth of a *Streptomyces* sp. strain of marine origin. (Color figure available online only.)

use such information for database searching, as described before for the standard case. We have also reported that these deconvoluted spectra of mixture components are fully searchable in our in-house spectral database providing faithful ranking rates comparable to those obtained when searching a standard proton spectrum of a pure compound [176].

Conclusions

We hope to have shown that efficient dereplication of MNPs requires access to both state-of-the-art analytical/spectroscopic instrumentation and to suitable databases. In-house chromatographic and spectral libraries play a major role in the dereplication workflow, thus being an extraordinary added value to any research group, institution, or company working on drug discovery from MNPs as has been illustrated with the dereplication workflow followed at Fundación MEDINA. The listed commercial databases and spectral libraries may look expensive in the short term but we suggest researchers in the field carefully evaluate their capabilities and usefulness since in most cases it is probable they will realize that it is worth the investment. In this sense we feel that the affirmation formulated 20 years ago by Corley and Durley [10] when they stated "The successful use of [commercial] databases for dereplication of natural products can result in considerable savings in time and money" keeps its validity today for MNP dereplication. On the other hand, the publicly available dereplication resources that we have highlighted deserve to be explored as a possible solution to any identification problem not only for being free of charge but also for providing, in some cases, an alternative or even more rapid route to dereplication. Ultimately, for achieving its maximum efficiency, the MNP dereplication toolbox should be employed smartly. To guarantee this, fortunately, the wisdom and experience of the MNP chemist still play an irreplaceable role.

Acknowledgements

Funding is acknowledged to the European Union's 7th Framework Programme via a Marie Curie Career Integration Grant (I.P.-V.) [PCIG-GA-2011–293762] and the PharmaSea project [Grant Agreement No 312184].

Merck & Co. Inc. (Rahway, NJ) is kindly acknowledged for the inheritance of its former MS Gold database and Deborah L. Zink, Dr. Claude Dufresne, and Dr. Jerrold Liesch are kindly acknowledged for developing it.

Prof. Stefan Günther and his team of the Pharmaceutical Bioinformatics group at the University of Freiburg are kindly acknowledged for developing the public StreptomeDB resource.

Dr. Kikuko Hayamizu and JEOL RESONANCE, Inc. are kindly acknowledged for making publicly available the CH-NMR-NP database.

Prof. Wolfgang Robien is kindly acknowledged for developing the public CSEARCH for PubChem resource.

Dr. Kristian F. Nielsen and Dr. Jens C. Frisvad are kindly acknowledged for making 277 compounds of their DTU (Technical Uni-

versity of Denmark) in-house mycotoxin-fungal secondary metabolite MS/HRMS library publicly accessible.
Prof. Peter Dorrestein and his team at USCD are kindly acknowledged for promoting the GNPS public MS/MS library initiative.

Conflict of Interest

The authors declare no conflict of interest.

References

1 *Kiuru P, D'Auria MV, Muller CD, Tammela P, Vuorela H, Yli-Kauhaluoma J.* Exploring marine resources for bioactive compounds. Planta Med 2014; 80: 1234–1246
2 *Molinski TF, Dalisay DS, Lievens SL, Saludes JP.* Drug development from marine natural products. Nat Rev Drug Discov 2009; 8: 69–85
3 *Harvey AL, Edrada-Ebel R, Quinn RJ.* The re-emergence of natural products for drug discovery in the genomics era. Nat Rev Drug Discov 2015; 14: 111–129
4 *Vanmiddlesworth F, Cannell RJP.* Dereplication and partial identification of natural products. In: Cannell RJP, editor. Natural products isolation. Totowa, NJ: Humana Press Inc.; 1998: 279–327
5 *Langlykke A.* Foreword. In: Bérdy J, editor. CRC handbook of antibiotic compounds, Vol. IV (Part 1). Boca Raton, FL: CRC Press; 1980
6 *Koehn FE, Carter GT.* The evolving role of natural products in drug discovery. Nat Rev Drug Discov 2005; 4: 206–220
7 *Blunt JW, Munro MHG.* Data, ^{1}H-NMR databases, data manipulation. Phytochem Rev 2013; 12: 435–447
8 Dictionary of Marine Natural Products (DMNP 2014). Chapman & Hall/CRC. Available at http://dmnp.chemnetbase.com/intro/index.jsp. Accessed October 20, 2015
9 MarinLit. A database of the marine natural products literature. RSC. Available at http://pubs.rsc.org/marinlit/. Accessed October 20, 2015
10 *Corley DG, Durley RC.* Strategies for database dereplication of natural products. J Nat Prod 1994; 57: 1484–1490
11 *Cordell GA, Shin YG.* Finding the needle in the haystack. The dereplication of natural product extracts. Pure Appl Chem 1999; 71: 1089–1094
12 *Bradshaw J, Butina D, Dunn AJ, Green RH, Hajek M, Jones MM, Lindon JC, Sidebottom PJ.* A rapid and facile method for the dereplication of purified natural products. J Nat Prod 2001; 64: 1541–1544
13 *Dinan L.* Dereplication and partial identification of natural products. In: Satyavit D, Sarker ZL, Gray AI, editors. Natural products isolation, 2nd edition. Totowa, NJ: Humana Press Inc.; 2006: 297–321
14 *Ito T, Masubuchi M.* Dereplication of microbial extracts and related analytical technologies. J Antibiot (Tokyo) 2014; 67: 353–360
15 *Halabalaki M, Vougogiannopoulou K, Mikros E, Skaltsounis AL.* Recent advances and new strategies in the NMR-based identification of natural products. Curr Opin Biotechnol 2014; 25: 1–7
16 *Wolfender JL, Marti G, Thomas A, Bertrand S.* Current approaches and challenges for the metabolite profiling of complex natural extracts. J Chromatogr A 2015; 1382: 136–164
17 *Nielsen KF, Larsen TO.* The importance of mass spectrometric dereplication in fungal secondary metabolite analysis. Front Microbiol 2015; 6: 71
18 *Williams RB, O'Neil-Johnson M, Williams AJ, Wheeler P, Pol R, Moser A.* Dereplication of natural products using minimal NMR data inputs. Org Biomol Chem 2015; 13: 9957–9962
19 *Gaudêncio SP, Pereira F.* Dereplication: racing to speed up the natural products discovery process. Nat Prod Rep 2015; 32: 779–810
20 *Blunt J, Munro M, Upjohn M.* The role of databases in marine natural products research. In: Fattorusso E, Gerwick WH, Taglialatela-Scafati O, editors. Handbook of marine natural products. Rotterdam: Springer Netherlands; 2012: 389–421
21 *Blunt JW, Munro MHG.* Is there an ideal database for natural products research? In: Osbourn A, Goss RJM, Carter GT, editors. Natural products: discourse, diversity, and design. Oxford, UK: John Wiley & Sons, Inc.; 2014: 413–431
22 *Lang G, Mayhudin NA, Mitova MI, Sun L, van der Sar S, Blunt JW, Cole AL, Ellis G, Laatsch H, Munro MH.* Evolving trends in the dereplication of natural product extracts: new methodology for rapid, small-scale investigation of natural product extracts. J Nat Prod 2008; 71: 1595–1599
23 *Mitova MI, Murphy AC, Lang G, Blunt JW, Cole AL, Ellis G, Munro MH.* Evolving trends in the dereplication of natural product extracts. 2. The isolation of chrysaibol, an antibiotic peptaibol from a New Zealand sample of the mycoparasitic fungus *Sepedonium chrysospermum*. J Nat Prod 2008; 71: 1600–1603
24 *Sultan S, Sun L, Blunt JW, Cole ALJ, Munro MHG, Ramasamy K, Weber JFF.* Evolving trends in the dereplication of natural product extracts. 3: further lasiodiplodins from *Lasiodiplodia theobromae*, an endophyte from *Mapania kurzii*. Tetrahedron Lett 2014; 55: 453–455
25 Chemical Abstracts Service (CAS) Registry, a division of the American Chemical Society. Available at https://www.cas.org/content/chemical-substances. Accessed October 20, 2015
26 SciFinder, a CAS solution. Available at http://www.cas.org/products/scifinder. Accessed October 20, 2015
27 REAXYS, Elsevier. Available at http://www.elsevier.com/solutions/reaxys. Accessed October 20, 2015
28 *Williams AJ.* Public chemical compound databases. Curr Opin Drug Discov Devel 2008; 11: 393–404
29 *Apodaca R.* Sixty-four free chemistry databases. Available at http://depth-first.com/articles/2011/10/12/sixty-four-free-chemistry-databases/. Accessed October 20, 2015
30 PubChem, NIH. Available at http://pubchem.ncbi.nlm.nih.gov. Accessed October 20, 2015
31 *Bolton EE, Wang Y, Thiessen PA, Bryant SH.* Chapter 12 – PubChem: integrated platform of small molecules and biological activities. Annu Rep Comput Chem 2008; 4: 217–241
32 ChemSpider, RSC. Available at http://www.chemspider.com/. Accessed October 20, 2015
33 *Gu J, Gui Y, Chen L, Yuan G, Lu HZ, Xu X.* Use of natural products as chemical library for drug discovery and network pharmacology. PLoS One 2013; 8: e62839
34 UNPD, Universal Natural Products Database of Peking University. Available at http://pkuxxj.pku.edu.cn/UNPD/. Accessed October 20, 2015
35 *Buckingham J, editor.* Dictionary of natural products on DVD v24: 1. Boca Raton: Chapman & Hall/CRC; 2015
36 Dictionary of Natural Products (DNP v24.1). Available at http://dnp.chemnetbase.com/intro/index.jsp. Accessed October 20, 2015
37 The DNP NMR Features database, available up to version 22.1, was prepared by John W Blunt, University of Canterbury, New Zealand, to whom enquiries about its preparation and use should be directed. john.blunt@canterbury.ac.nz
38 *Laatsch H.* AntiBase 2014: the natural compound identifier. Available at http://eu.wiley.com/WileyCDA/WileyTitle/productCd-3527338411.html. Accessed October 20, 2015
39 SpecInfo, Wiley Online Library Spectroscopy. Available at http://www.wiley-vch.de/stmdata/specinfo.php. Accessed October 20, 2015
40 The AntiMarin database, a combination database formed from AntiBase and MarinLit, was prepared by John W Blunt, University of Canterbury, New Zealand, to whom enquiries about its preparation and use should be directed. john.blunt@canterbury.ac.nz
41 *Lucas X, Senger C, Erxleben A, Grüning BA, Döring K, Mosch J, Flemming S, Günther S.* StreptomeDB: a resource for natural compounds isolated from *Streptomyces* species. Nucleic Acids Res 2013; 41: D1130–D1136
42 StreptomeDB 2.0. Available at http://www.pharmaceutical-bioinformatics.de/streptomedb/. Accessed October 20, 2015
43 *Allen F, Greiner R, Wishart D.* Competitive fragmentation modeling of ESI-MS/MS spectra for putative metabolite identification. Metabolomics 2014; 11: 98–110
44 Marvin Applets. Available at https://www.chemaxon.com/marvin/. Accessed October 20, 2015
45 ACD/NMR Predictors (ACD/Labs). Available at http://www.acdlabs.com/products/adh/nmr/nmr_pred/. Accessed October 20, 2015
46 ACD/Structure Elucidator Suite (ACD/Labs). Available at http://www.acdlabs.com/products/com_iden/elucidation/struc_eluc/. Accessed October 20, 2015
47 MICRONMR (Shanghai Micronmr Infor Technology Co., Ltd.) Available at http://www.nmrdata.com: 90/masterInfor_introduce.aspx. Accessed October 20, 2015
48 *Yang Z, Wu Y, Zhou H, Cao X, Jiang X, Wang K, Wu S.* A novel strategy for screening new natural products by a combination of reversed-phase liquid chromatography fractionation and ^{13}C NMR pattern recognition: the discovery of new anti-cancer flavone dimers from *Dysosma versipellis* (Hance). RSC Adv 2015; 5: 77553–77564

49 *Hayamizu KY, Asakura K, Kurimoto T.* An open access NMR database for organic natural products "CH-NMR-NP". Prague, Czech Republic: EUROMAR; 2015
50 *Hayamizu KY, Asakura K, Kurimoto T.* An open access NMR database for organic natural products "CH-NMR-NP". 57th Experimental Nuclear Magnetic Resonance Conference, Pittsburgh, PA; 2015
51 *Hayamizu K.* [On an NMR database for natural products "CH-NMR-NP"]. Kagaku to Seibutsu 2011; 49: 250–255
52 *Hayamizu K.* Natural Product NMR-DB "CH-NMR-NP". Available at https://www.j-resonance.com/en/nmrdb/. Accessed October 20, 2015
53 *López-Pérez JL, Therón R, del Olmo E, Díaz D.* NAPROC-13: a database for the dereplication of natural product mixtures in bioassay-guided protocols. Bioinformatics 2007; 23: 3256–3257
54 NAPROC-13. Available at http://c13.usal.es/c13/usuario/views/inicio.jsp?lang=es&country=ES. Accessed October 20, 2015
55 *Mihaleva VV, te Beek TA, van Zimmeren F, Moco S, Laatikainen R, Niemitz M, Korhonen SP, van Driel MA, Vervoort J.* MetIDB: A publicly accessible database of predicted and experimental ^{1}H NMR spectra of flavonoids. Anal Chem 2013; 85: 8700–8707
56 MetIDB. Available at http://metidb.org/home. Accessed October 20, 2015
57 *Fischedick JT, Johnson SR, Ketchum RE, Croteau RB, Lange BM.* NMR spectroscopic search module for Spektraris, an online resource for plant natural product identification – Taxane diterpenoids from *Taxus × media* cell suspension cultures as a case study. Phytochemistry 2015; 113: 87–95
58 Spektraris NMR. Available at http://langelabtools.wsu.edu/nmr/. Accessed October 20, 2015
59 *Steinbeck C, Kuhn S.* NMRShiftDB – compound identification and structure elucidation support through a free community-built web database. Phytochemistry 2004; 65: 2711–2717
60 NMRShiftDB. Available at http://nmrshiftdb.nmr.uni-koeln.de/. Accessed October 20, 2015
61 *Biological Magnetic Resonance Databank (BMRB).* Available at http://www.bmrb.wisc.edu/metabolomics/. Accessed October 20, 2015
62 *Ulrich EL, Akutsu H, Doreleijers JF, Harano Y, Ioannidis YE, Lin J, Livny M, Mading S, Maziuk D, Miller Z, Nakatani E, Schulte CF, Tolmie DE, Kent Wenger R, Yao H, Markley JL.* BioMagResBank. Nucleic Acids Res 2008; 36: D402–D408
63 *Wishart DS, Jewison T, Guo AC, Wilson M, Knox C, Liu Y, Djoumbou Y, Mandal R, Aziat F, Dong E, Bouatra S, Sinelnikov I, Arndt D, Xia J, Liu P, Yallou F, Bjorndahl T, Perez-Pineiro R, Eisner R, Allen F, Neveu V, Greiner R, Scalbert A.* HMDB 3.0–The Human Metabolome Database in 2013. Nucleic Acids Res 2013; 41: D801–D807
64 The Human Metabolome Database (HMDB). Available at http://www.hmdb.ca/. Accessed October 20, 2015
65 *Cui Q, Lewis IA, Hegeman AD, Anderson ME, Li J, Schulte CF, Westler WM, Eghbalnia HR, Sussman MR, Markley JL.* Metabolite identification via the Madison metabolomics consortium database. Nat Biotechnol 2008; 26: 162–164
66 Madison Metabolomics Consortium Database (MMCD). Available at http://mmcd.nmrfam.wisc.edu/. Accessed October 20, 2015
67 *Bingol K, Zhang F, Bruschweiler-Li L, Brüschweiler R.* TOCCATA: a customized carbon total correlation spectroscopy NMR metabolomics database. Anal Chem 2012; 84: 9395–9401
68 COLMAR ^{13}C-TOCCATA: a carbon TOCSY NMR metabolomics database. Available at http://spin.ccic.ohio-state.edu/index.php/toccata/index. Accessed October 20, 2015
69 *Bingol K, Bruschweiler-Li L, Li DW, Brüschweiler R.* Customized metabolomics database for the analysis of NMR ^{1}H-^{1}H TOCSY and ^{13}C-^{1}H HSQC-TOCSY spectra of complex mixtures. Anal Chem 2014; 86: 5494–5501
70 COLMAR ^{1}H(^{13}C)-TOCCATA: customized metabolomics database for the analysis of NMR ^{1}H-^{1}H TOCSY and ^{13}C-^{1}H HSQC-TOCSY spectra of complex mixtures. Available at http://spin.ccic.ohio-state.edu/index.php/toccata2/index. Accessed October 20, 2015
71 *Bingol K, Li DW, Bruschweiler-Li L, Cabrera OA, Megraw T, Zhang F, Brüschweiler R.* Unified and isomer-specific NMR metabolomics database for the accurate analysis of ^{13}C-^{1}H HSQC spectra. ACS Chem Biol 2015; 10: 452–459
72 COLMAR ^{13}C-^{1}H HSQC query. Available at http://spin.ccic.ohio-state.edu/index.php/hsqc/index. Accessed January 20, 2016
73 *Johnson SR, Lange BM.* Open-access metabolomics databases for natural product research: present capabilities and future potential. Front Bioeng Biotechnol 2015; 3: 22
74 *Horai H, Arita M, Kanaya S, Nihei Y, Ikeda T, Suwa K, Ojima Y, Tanaka K, Tanaka S, Aoshima K, Oda Y, Kakazu Y, Kusano M, Tohge T, Matsuda F, Sawada Y, Hirai MY, Nakanishi H, Ikeda K, Akimoto N, Maoka T, Takahashi H, Ara T, Sakurai N, Suzuki H, Shibata D, Neumann S, Iida T, Funatsu K, Matsuura F, Soga T, Taguchi R, Saito K, Nishioka T.* MassBank: a public repository for sharing mass spectral data for life sciences. J Mass Spectrom 2010; 45: 703–714
75 MassBank. Available at http://www.massbank.jp/en/database.html. Accessed October 20, 2015
76 *Smith CA, O'Maille G, Want EJ, Qin C, Trauger SA, Brandon TR, Custodio DE, Abagyan R, Siuzdak G.* METLIN: a metabolite mass spectral database. Ther Drug Monit 2005; 27: 747–751
77 METLIN. Available at https://metlin.scripps.edu/index.php. Accessed October 20, 2015
78 *Nielsen KF, Frisvad JC.* DTU mycotoxin-fungal secondary metabolite MS/HRMS library. Available at http://www.bio.dtu.dk/english/Research/Platforms/Metabolom/MSMSLib. Accessed October 20, 2015
79 GNPS Public Spectral Libraries. Available at http://gnps.ucsd.edu/ProteoSAFe/libraries.jsp. Accessed October 20, 2015
80 *Strege MA.* Hydrophilic interaction chromatography-electrospray mass spectrometry analysis of polar compounds for natural product drug discovery. Anal Chem 1998; 70: 2439–2445
81 *Frisvad JC, Thrane U.* Standardized high-performance liquid chromatography of 182 mycotoxins and other fungal metabolites based on alkylphenone retention indices and UV-VIS spectra (diodearray detection). J Chromatogr A 1987; 404: 195–214
82 *Nielsen KF, Smedsgaard J.* Fungal metabolite screening: database of 474 mycotoxins and fungal metabolites for dereplication by standardised liquid chromatography-UV-mass spectrometry methodology. J Chromatogr A 2003; 1002: 111–136
83 *Hill DW, Kelley TR, Laugner KJ, Miller KW.* Determination of mycotoxins by gradient high-performance liquid chromatography using an alkylphenone retention index system. Anal Chem 1984; 56: 2576–2579
84 *Stanstrup J, Neumann S, Vrhovšek U.* PredRet: prediction of retention time by direct mapping between multiple chromatographic systems. Anal Chem 2015; 87: 9421–9428
85 *Boswell PG, Schellenberg JR, Carr PW, Cohen JD, Hegeman AD.* 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 2011; 1218: 6732–6741
86 *Abate-Pella D, Freund DM, Ma Y, Simón-Manso Y, Hollender J, Broeckling CD, Huhman DV, Krokhin OV, Stoll DR, Hegeman AD, Kind T, Fiehn O, Schymanski EL, Prenni JE, Sumner LW, Boswell PG.* Retention projection enables accurate calculation of liquid chromatographic retention times across labs and methods. J Chromatogr A 2015; 1412: 43–51
87 *Cao M, Fraser K, Huege J, Featonby T, Rasmussen S, Jones C.* Predicting retention time in hydrophilic interaction liquid chromatography mass spectrometry and its use for peak annotation in metabolomics. Metabolomics 2015; 11: 696–706
88 *Eugster PJ, Boccard J, Debrus B, Bréant L, Wolfender JL, Martel S, Carrupt PA.* Retention time prediction for dereplication of natural products (CxHyOz) in LC-MS metabolite profiling. Phytochemistry 2014; 108: 196–207
89 *Creek DJ, Jankevics A, Breitling R, Watson DG, Barrett MP, Burgess KEV.* Toward global metabolomics analysis with hydrophilic interaction liquid chromatography-mass spectrometry: improved metabolite identification by retention time prediction. Anal Chem 2011; 83: 8703–8710
90 *Seger C, Sturm S, Stuppner H.* Mass spectrometry and NMR spectroscopy: modern high-end detectors for high resolution separation techniques-state of the art in natural product HPLC-MS, HPLC-NMR, and CE-MS hyphenations. Nat Prod Rep 2013; 30: 970–987
91 *Wolfender JL.* HPLC in natural product analysis: the detection issue. Planta Med 2009; 75: 719–734
92 *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 2013; 30: 546–564
93 *Potterat O, Hamburger M.* Combined use of extract libraries and HPLC-based activity profiling for lead discovery: potential, challenges, and practical considerations. Planta Med 2014; 80: 1171–1181
94 *Lang G, Mitova MI, Ellis G, van der Sar S, Phipps RK, Blunt JW, Cummings NJ, Cole ALJ, Munro MHG.* Bioactivity profiling using HPLC/microtiterplate analysis: application to a New Zealand marine alga-derived fungus, *Gliocladium* sp. J Nat Prod 2006; 69: 621–624

95 *Johnson TA, Sohn J, Inman WD, Estee SA, Loveridge ST, Vervoort HC, Tenney K, Liu J, Ang KKH, Ratnam J, Bray WM, Gassner NC, Shen YY, Lokey RS, McKerrow JH, Boundy-Mills K, Nukanto A, Kanti A, Julistiono H, Kardono LBS, Bjeldanes LF, Crews P.* Natural product libraries to accelerate the high-throughput discovery of therapeutic leads. J Nat Prod 2011; 74: 2545–2555

96 *Bugni TS, Richards B, Bhoite L, Cimbora D, Harper MK, Ireland CM.* Marine natural product libraries for high-throughput screening and rapid drug discovery. J Nat Prod 2008; 71: 1095–1098

97 *Bohni N, Cordero-Maldonado ML, Maes J, Siverio-Mota D, Marcourt L, Munck S, Kamuhabwa AR, Moshi MJ, Esguerra CV, de Witte PAM, Crawford AD, Wolfender JL.* Integration of microfractionation, qNMR and zebrafish screening for the *in vivo* bioassay-guided isolation and quantitative bioactivity analysis of natural products. PLoS One 2013; 8: e64006

98 *Challal S, Buenafe OEM, Queiroz EF, Maljevic S, Marcourt L, Bock M, Kloeti W, Dayrit FM, Harvey AL, Lerche H, Esguerra CV, De Witte PAM, Wolfender JL, Crawford AD.* Zebrafish bioassay-guided microfractionation identifies anticonvulsant steroid glycosides from the Philippine medicinal plant Solanum torvum. ACS Chem Neurosci 2014; 5: 993–1004

99 *Challal S, Bohni N, Buenafe OE, Esguerra CV, De Witte PAM, Wolfender JL, Crawford AD.* Zebrafish bioassay-guided microfractionation for the rapid *in vivo* identification of pharmacologically active natural products. Chimia (Aarau) 2012; 66: 229–232

100 *Nielsen KF, Månsson M, Rank C, Frisvad JC, Larsen TO.* Dereplication of microbial natural products by LC-DAD-TOFMS. J Nat Prod 2011; 74: 2338–2348

101 *El-Elimat T, Figueroa M, Ehrmann BM, Cech NB, Pearce CJ, Oberlies NH.* High-resolution MS, MS/MS, and UV database of fungal secondary metabolites as a dereplication protocol for bioactive natural products. J Nat Prod 2013; 76: 1709–1716

102 *Zink D, Dufresne C, Liesch J, Martín J.* Automated LC-MS analysis of natural products: extraction of UV, MS and retention time data for component identification and characterization. Proceedings of the 50th ASMS Conference on Mass Spectrometry and Allied Topics, Orlando, FL; 2002

103 *Zink D, Dufresne C, Liesch J, Martín J.* Identification/dereplication of natural products by LC-UV-MS. Spectral search parameters. Small Molecule Science Conference (COSMOS), Bristol, RI; 2005

104 *Hansen ME, Smedsgaard J, Larsen TO.* X-hitting: an algorithm for novelty detection and dereplication by UV spectra of complex mixtures of natural products. Anal Chem 2005; 77: 6805–6817

105 *Larsen TO, Hansen MAE.* Dereplication and discovery of natural products by UV spectroscopy. In: Colegate SM, Molyneux RJ, editors. Bioactive natural products. Detection, isolation and structural determination, 2nd edition. Boca Raton, FL: CRC Press; 2008: 221–244

106 *Wehrens R, Carvalho E, Fraser PD.* Metabolite profiling in LC-DAD using multivariate curve resolution: the alsace package for R. Metabolomics 2014; 11: 143–154

107 *Larsen TO, Petersen BO, Duus JØ, Sørensen D, Frisvad JC, Hansen ME.* Discovery of new natural products by application of X-hitting, a novel algorithm for automated comparison of full UV spectra, combined with structural determination by NMR spectroscopy. J Nat Prod 2005; 68: 871–874

108 *Kind T, Fiehn O.* Metabolomic database annotations via query of elemental compositions: mass accuracy is insufficient even at less than 1 ppm. BMC Bioinformatics 2006; 7: 234

109 *Fredenhagen A, Derrien C, Gassmann E.* An MS/MS library on an ion-trap instrument for efficient dereplication of natural products. Different fragmentation patterns for $[M + H]^+$ and $[M + Na]^+$ ions. J Nat Prod 2005; 68: 385–391

110 *Klitgaard A, Iversen A, Andersen MR, Larsen TO, Frisvad JC, Nielsen KF.* Aggressive dereplication using UHPLC-DAD-QTOF: screening extracts for up to 3000 fungal secondary metabolites. Anal Bioanal Chem 2014; 406: 1933–1943

111 *Kildgaard S, Mansson M, Dosen I, Klitgaard A, Frisvad JC, Larsen TO, Nielsen KF.* Accurate dereplication of bioactive secondary metabolites from marine-derived fungi by UHPLC-DAD-QTOFMS and a MS/HRMS library. Mar Drugs 2014; 12: 3681–3705

112 *Scheubert K, Hufsky F, Böcker S.* Computational mass spectrometry for small molecules. J Cheminform 2013; 5: 12

113 *Hufsky F, Scheubert K, Böcker S.* Computational mass spectrometry for small-molecule fragmentation. Trends Anal Chem 2014; 53: 41–48

114 *Vaniya A, Fiehn O.* Using fragmentation trees and mass spectral trees for identifying unknown compounds in metabolomics. Trends Anal Chem 2015; 69: 52–61

115 Mass Frontier Spectral Interpretation Software (ThermoScientific). Available at http://www.thermoscientific.com/en/product/mass-frontier-7-0-spectral-interpretation-software.html. Accessed October 20, 2015

116 ACD/MS Fragmenter (ACD/Labs). Available at http://www.acdlabs.com/products/adh/ms/ms_frag/. Accessed October 20, 2015

117 MetFrag. Available at http://msbi.ipb-halle.de/MetFrag/. Accessed October 20, 2015

118 *Wolf S, Schmidt S, Müller-Hannemann M, Neumann S. In silico* fragmentation for computer assisted identification of metabolite mass spectra. BMC Bioinformatics 2010; 11: 148

119 FingerID. Available at http://research.ics.aalto.fi/kepaco/fingerid/index.html. Accessed October 20, 2015

120 *Heinonen M, Shen H, Zamboni N, Rousu J.* Metabolite identification and molecular fingerprint prediction through machine learning. Bioinformatics 2012; 28: 2333–2341

121 *Allen F, Pon A, Wilson M, Greiner R, Wishart D.* CFM-ID: a web server for annotation, spectrum prediction and metabolite identification from tandem mass spectra. Nucleic Acids Res 2014; 42: W94–W99

122 CFM-ID. Available at http://cfmid.wishartlab.com/. Accessed October 20, 2015

123 MAGMa. Available at http://www.emetabolomics.org/magma. Accessed October 20, 2015

124 *Ridder L, van der Hooft JJJ, Verhoeven S, de Vos RCH, Bino RJ, Vervoort J.* Automatic chemical structure annotation of an $LC\text{-}MS^n$ based metabolic profile from green tea. Anal Chem 2013; 85: 6033–6040

125 *Ridder L, Van Der Hooft JJJ, Verhoeven S, De Vos RCH, Van Schaik R, Vervoort J.* Substructure-based annotation of high-resolution multistage MS^n spectral trees. Rapid Commun Mass Spectrom 2012; 26: 2461–2471

126 *Ridder L, van der Hooft JJJ, Verhoeven S, de Vos RCH, Vervoort J, Bino RJ. In silico* prediction and automatic $LC\text{-}MS^n$ annotation of green tea metabolites in urine. Anal Chem 2014; 86: 4767–4774

127 CSI : FingerID. Available at http://www.csi-fingerid.org/. Accessed October 20, 2015

128 *Dührkop K, Shen H, Meusel M, Rousu J, Böcker S.* Searching molecular structure databases with tandem mass spectra using CSI : FingerID. Proc Natl Acad Sci U S A 2015; 112: 12580–12585

129 *Yang JY, Sanchez LM, Rath CM, Liu X, Boudreau PD, Bruns N, Glukhov E, Wodtke A, De Felicio R, Fenner A, Wong WR, Linington RG, Zhang L, Debonsi HM, Gerwick WH, Dorrestein PC.* Molecular networking as a dereplication strategy. J Nat Prod 2013; 76: 1686–1699

130 *Jaroszewski JW.* Hyphenated NMR methods in natural products research, part 1: direct hyphenation. Planta Med 2005; 71: 691–700

131 *Jaroszewski JW.* Hyphenated NMR methods in natural products research, Part 2: HPLC-SPE-NMR and other new trends in NMR hyphenation. Planta Med 2005; 71: 795–802

132 *Seger C, Godejohann M, Tseng LH, Spraul M, Girtler A, Sturm S, Stuppner H.* LC-DAD-MS/SPE-NMR hyphenation. A tool for the analysis of pharmaceutically used plant extracts: identification of isobaric iridoid glycoside regioisomers from Harpagophytum procumbens. Anal Chem 2005; 77: 878–885

133 *Jansma A, Chuan T, Albrecht RW, Olson DL, Peck TL, Geierstanger BH.* Automated microflow NMR: routine analysis of five-microliter samples. Anal Chem 2005; 77: 6509–6515

134 *Lambert M, Wolfender JL, Stærk D, Christensen SB, Hostettmann K, Jaroszewski JW.* Identification of natural products using HPLC-SPE combined with CapNMR. Anal Chem 2007; 79: 727–735

135 *Schroeder FC, Gronquist M.* Extending the scope of NMR spectroscopy with microcoil probes. Angew Chem Int Ed Engl 2006; 45: 7122–7131

136 *Hu J, Eldridge GR, Yu Y, O'Neil-Johnson M.* High-throughput natural product chemistry methods and the application of the capillary NMR probe. Prog Chem 2008; 20: 429–440

137 *Hilton BD, Martin GE.* Investigation of the experimental limits of small-sample heteronuclear 2D NMR. J Nat Prod 2010; 73: 1465–1469

138 *Molinski TF.* NMR of natural products at the 'nanomole-scale'. Nat Prod Rep 2010; 27: 321–329

139 ACD/Spectrus DB (ACD/Labs). Available at http://www.acdlabs.com/products/spectrus/db/. Accessed October 20, 2015

140 MNova DB (Mestrelab Research). Available at http://mestrelab.com/software/mnova/db/. Accessed October 20, 2015

141 AMIX (Bruker Biospin). Available at https://www.bruker.com/products/mr/nmr/nmr-software/software/amix/overview.html. Accessed October 20, 2015

142 *Johansen KT, Wubshet SG, Nyberg NT.* HPLC-NMR revisited: using time-slice high-performance liquid chromatography-solid-phase extraction-nuclear magnetic resonance with database-assisted dereplication. Anal Chem 2013; 85: 3183–3189

143 ACD/Labs NMR Databases. Available at http://www.acdlabs.com/products/adh/spectrusprocessor/wiley_nmr/. Accessed October 20, 2015

144 KnowItAll NMR Spectral Library (Bio-Rad). Available at http://www.bio-rad.com/es-es/product/nmr-spectral-databases. Accessed October 20, 2015

145 *Robien W.* CSEARCH. Available at http://nmrpredict.orc.univie.ac.at/. Accessed October 20, 2015

146 *Hubert J, Nuzillard JM, Purson S, Hamzaoui M, Borie N, Reynaud R, Renault JH.* Identification of natural metabolites in mixture: a pattern recognition strategy based on ^{13}C NMR. Anal Chem 2014; 86: 2955–2962

147 *Tsipouras A, Ondeyka J, Dufresne C, Lee S, Salituro G, Tsou N, Goetz M, Singh SB, Kearsley SK.* Using similarity searches over databases of estimated ^{13}C NMR spectra for structure identification of natural product compounds. Anal Chim Acta 1995; 316: 161–171

148 *Smith SK, Cobleigh J, Svetnik V.* Evaluation of a ^{1}H-^{13}C NMR spectral library. J Chem Inf Comput Sci 2001; 41: 1463–1469

149 NMRanalyst. Available at http://www.sciencesoft.net/NMRanalyst.html. Accessed October 20, 2015

150 *Dunkel R, Wu X.* Identification of organic molecules from a structure database using proton and carbon NMR analysis results. J Magn Reson 2007; 188: 97–110

151 *Bremser W, Wagner H, Franke B.* Fast searching for identical ^{13}C NMR spectra via inverted files. Org Magn Reson 1981; 15: 178–187

152 *Robien W.* CSEARCH for PubChem. Available at http://nmrpredict.orc.univie.ac.at/similar/eval.php. Accessed October 20, 2015

153 Fundación MEDINA. Centro de Excelencia en Investigación de Medicamentos Innovadores en Andalucía. Available at http://www.medinadiscovery.com/. Accessed October 20, 2015

154 *Genilloud O, González I, Salazar O, Martín J, Tormo JR, Vicente F.* Current approaches to exploit actinomycetes as a source of novel natural products. J Ind Microbiol Biotechnol 2011; 38: 375–389

155 *Genilloud O, Vicente F.* Strategies to discover novel antimicrobials to cope with emerging medical needs. In: Marinelli F, Genilloud O, editors. Antimicrobials: new and old molecules in the fight against multi-resistant bacteria. Berlin: Springer-Verlag; 2014: 327–360

156 *Genilloud O.* The re-emerging role of microbial natural products in antibiotic discovery. Antonie Van Leeuwenhoek 2014; 106: 173–188

157 *Annang F, Pérez-Moreno G, García-Hernández R, Cordon-Obras C, Martín J, Tormo JR, Rodríguez L, De Pedro N, Gómez-Pérez V, Valente M, Reyes F, Genilloud O, Vicente F, Castanys S, Ruiz-Pérez LM, Navarro M, Gamarro F, González-Pacanowska D.* High-throughput screening platform for natural product-based drug discovery against 3 neglected tropical diseases: human African trypanosomiasis, leishmaniasis, and chagas disease. J Biomol Screen 2015; 20: 82–91

158 *Martín J, Crespo G, González-Menéndez V, Pérez-Moreno G, Sánchez-Carrasco P, Pérez-Victoria I, Ruiz-Pérez LM, González-Pacanowska D, Vicente F, Genilloud O, Bills GF, Reyes F.* MDN-0104, an antiplasmodial betaine lipid from Heterospora chenopodii. J Nat Prod 2014; 77: 2118–2123

159 *Cautain B, De Pedro N, Garzón VM, De Escalona MM, González Menéndez V, Tormo JR, Martin J, El Aouad N, Reyes F, Asensio F, Genilloud O, Vicente F, Link W.* High-content screening of natural products reveals novel nuclear export inhibitors. J Biomol Screen 2014; 19: 57–65

160 *Monteiro MC, De La Cruz M, Cantizani J, Moreno C, Tormo JR, Mellado E, De Lucas JR, Asensio F, Valiante V, Brakhage AA, Latgé JP, Genilloud O, Vicente F.* A new approach to drug discovery: high-throughput screening of microbial natural extracts against *Aspergillus fumigatus* using resazurin. J Biomol Screen 2012; 17: 542–549

161 PharmaSea. Available at http://www.pharma-sea.eu/pharmasea.html. Accessed October 20, 2015

162 *Stein SE.* An integrated method for spectrum extraction and compound identification from gas chromatography/mass spectrometry data. J Am Soc Mass Spectrom 1999; 10: 770–781

163 AMDIS. Available at http://www.amdis.net/. Accessed October 20, 2015

164 *Lacret R, Oves-Costales D, Gómez C, Díaz C, De La Cruz M, Pérez-Victoria I, Vicente F, Genilloud O, Reyes F.* New ikarugamycin derivatives with antifungal and antibacterial properties from Streptomyces zhaozhouensis. Mar Drugs 2015; 13: 128–140

165 *Tong H, Bell D, Tabei K, Siegel MM.* Automated data massaging, interpretation, and e-mailing modules for high throughput open access mass spectrometry. J Am Soc Mass Spectrom 1999; 10: 1174–1187

166 SmartFormula 3D (Bruker Daltonics). Available at https://www.bruker.com/fileadmin/user_upload/8-PDF-Docs/Separations_MassSpectrometry/Literature/literature/TechNotes/TN-26_smartformula3D_12-2014_eBook.pdf. Accessed October 20, 2015

167 *Motohashi K, Takagi M, Shin-Ya K.* Tetrapeptides possessing a unique skeleton, JBIR-34 and JBIR-35, isolated from a sponge-derived actinomycete, *Streptomyces* sp. Sp080513GE-23. J Nat Prod 2010; 73: 226–228

168 *Martín J, Pérez-Victoria I, González V, de Pedro N, Vicente F, Bills G, Reyes F.* Applying LC-MS de-replication strategies for the discovery of new natural products. Planta Med 2012; 78: PI77

169 *Takada K, Ninomiya A, Naruse M, Sun Y, Miyazaki M, Nogi Y, Okada S, Matsunaga S.* Surugamides A–E, cyclic octapeptides with four D-amino acid residues, from a marine *Streptomyces* sp.: LC-MS-aided inspection of partial hydrolysates for the distinction of D- and L-amino acid residues in the sequence. J Org Chem 2013; 78: 6746–6750

170 *Pesic A, Baumann HI, Kleinschmidt K, Ensle P, Wiese J, Süssmuth RD, Imhoff JF.* Champacyclin, a new cyclic octapeptide from *Streptomyces* strain C42 isolated from the Baltic Sea. Mar Drugs 2013; 11: 4834–4857

171 *Schmidt JS, Lauridsen MB, Dragsted LO, Nielsen J, Staerk D.* Development of a bioassay-coupled HPLC-SPE-ttNMR platform for identification of α-glucosidase inhibitors in apple peel (*Malus* × *domestica* Borkh.). Food Chem 2012; 135: 1692–1699

172 *Claridge TDW.* High-resolution NMR techniques in organic chemistry, 2nd edition. Oxford, UK: Elsevier; 2009: 303–334

173 *Antalek B.* Using pulsed gradient spin echo NMR for chemical mixture analysis: How to obtain optimum results. Concepts Magn Reson 2002; 14: 225–258

174 *Pérez-Victoria I.* DOSYMNPs: Novel applications of diffusion NMR spectroscopy in microbial natural products research. Available at http://cordis.europa.eu/project/rcn/99978_en.html. Accessed October 20, 2015

175 *Evans R, Deng Z, Rogerson AK, McLachlan AS, Richards JJ, Nilsson M, Morris GA.* Quantitative interpretation of diffusion-ordered NMR spectra: can we rationalize small molecule diffusion coefficients? Angew Chem Int Ed Engl 2013; 52: 3199–3202

176 *Pérez-Victoria I, Crespo G, Reyes F.* Dereplication of natural products in mixtures using PFG diffusion NMR combined with an in-house ^{1}H NMR spectra database. Small Molecule NMR Conference (SMASH), Santiago de Compostela; 2013