doi:10.2533/chimia.2007.332 CHIMIA **2007**, *61*, No. 6

Chimia 61 (2007) 332–338

ISSN 0009–4293

Accelerated Dereplication of Natural Products, Supported by Reference Libraries

Jens Bitzer[a], Bärbel Köpcke[a], Marc Stadler[a], Veronika Hellwig[b], Yu-Ming Ju[c], Stephan Seip[d], and Thomas Henkel*[a]

Abstract: Natural products are an indispensable source for drug discovery. The major challenge for exploiting this evolutionary optimized pool of potential lead structures is the fast and reliable recognition of known compounds, *i.e.* dereplication. This task is essential for the discovery process in high-throughput screening scenarios, since it allows the focus to be placed on novel chemical structures at an early stage. Furthermore, information on identified compounds will help to rationalize observed bioactivities. This article describes an effective, library-supported strategy for the dereplication of crude extracts and pre-fractionated samples, using an HPLC-based multidetector platform and NMR techniques, respectively.

Keywords: Dereplication · Multidetector HPLC · Natural products · NMR · Secondary metabolites · Tandem mass spectrometry

1. Introduction

Natural sources provide a fascinating variety of structurally distinct and biologically active secondary metabolites. These natural products (NPs) proved to be an indispensable source for lead compounds for the pharmaceutical[1] and agrochemical[2] industries. Many NPs are in use as essential tools for biochemistry and fundamental biomolecular research.[3] Their unique structural features comprise a chemical space that cannot be covered by synthetic chemistry,[4] thus providing benefit for every discovery process.[5] However, it is estimated that clearly less than 10% of the world's natural chemodiversity has been explored yet.[6] Regarding the world of micro-organisms, where modern biomolecular techniques gave insight into an unexpected and still largely untapped phylogenetic diversity both in terrestrial and marine environments, this number may be even significantly smaller.[7] Moreover, genome sequencing of the antibiotic-producing actinomycete *Streptomyces coelicolor* revealed the presence of more than 20 secondary metabolite gene clusters. However, fewer than half of these compound classes have hitherto been identified from this strain,[8] implying that novel metabolites can still be obtained from organisms that have already been screened for decades. Concise variation of fermentation parameters and genetic manipulation approaches may help in the future to further explore their non-exhausted biosynthetic potential.[9] An unexploited biosynthetic potential at least equivalent to filamentous bacteria can be found in fungi, *e.g.* genome mining revealed over 40 genes encoding for secondary metabolites in *Aspergillus nidulans*.[10]

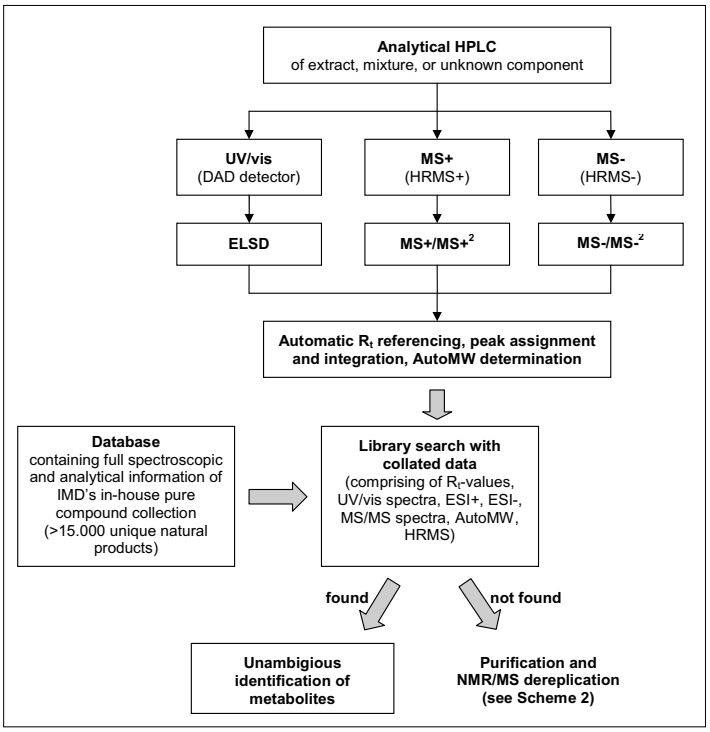
Irrespective whether a chemical or a biological screening approach is chosen, the crucial step for finding new lead structures is the fast and reliable recognition of already known compounds. This process is called dereplication. It allows new chemical structures to be focused on at a very early stage of the discovery process, thus avoiding the redundant re-isolation of known metabolites that may no longer be patentable. Additionally, the dereplication process will rapidly provide access to all information that is linked with the identified component, *e.g.* bioactivities, alternative sources, and applications.

Today, dereplication of natural products is usually achieved by using high-performance liquid chromatography (HPLC) coupled with UV/Vis spectroscopy and mass spectrometry (MS).[11] More sophisticated approaches have also been described applying MS/MS (MS^n) techniques, taking advantage of the additional information of fragmentation patterns, and thus allowing for recognition of substructures.[11,12] Last not least, HPLC-NMR coupling is discussed as an alternative technique, and its utility has improved considerably by the availability of HPLC-SPE-NMR.[13] But as this technique requires rather expensive equipment, it has not yet become a routine methodology. Moreover, the sample throughput is rather low compared to conventional LC-UV/MS techniques. As soon as the number of samples used for screening increases, *e.g.* to meet the requirements of high-through-

*Correspondence: Dr. T. Henkel[a]
Tel.: +49 231 9742 6060
Fax: +49 231 9742 6061
E-Mail: thomas.henkel@intermed-discovery.com
[a]InterMed Discovery GmbH
Otto-Hahn-Straße 15
D-44227 Dortmund, Germany
[b]University of Lüneburg
Institute for Ecology and Environmental Chemistry and GKSS-Research Centre, Institute for Coastal Research
D-21502 Geesthacht, Germany
[c]Academia Sinica
Institute of Plant and Microbial Biology
Nankang, Taipei 11529, Taiwan
[d]Bayer HealthCare AG
D-54368 Leverkusen, Germany

put screening (HTS), a large number of hits will need to be evaluated within a narrow timeframe and usually with restricted capacities for re-fermentation, large-scale isolation, and other laborious procedures. Sound criteria for early-stage pre-selection of primary screening hits need to be established. At InterMed Discovery (IMD), databases on in-house screening data are linked to the dereplication process, hence various kinds of information related to the identified components (*e.g.* bioactivities including toxicity, alternative sources, and applications) can be used to facilitate selection of hits for intensified evaluation (large-scale fermentation and re-isolation).[14]

Here we describe a two-level approach elaborated at IMD for effective and accelerated dereplication of natural products. The first level comprises an analytical HPLC system coupled with multiple detectors. The collated and corrected data of all detectors are used for a single search in a reference library. If the metabolite of interest cannot be identified by the first level, the second level which we call NMR/MS dereplication involves its (partial) purification and basic NMR analysis. By simply recording ^{1}H and $^{1}H,^{13}C$-HSQC spectra followed by minimal interpretation, a software tool checks for structural identity within a database containing all literature-known NPs. Noteworthy, this step is independent from the physical availability of the reference compound. For any given known structure, the subsequent verification of its identity is usually a straightforward process. If the compound has not been identified in the second step, it most probably constitutes a novel metabolite awaiting its first-time structural elucidation.



Scheme 1. Accelerated library-supported dereplication of metabolites in extracts or mixtures using a multidetector HPLC platform

2. Accelerated Recognition of Known Secondary Metabolites by a Multidetector HPLC Platform Supported by Reference Libraries

The first level is built up by an analytical HPLC system coupled with a diode array detector (DAD), an evaporative light scattering detector (ELSD), and mass spectrometers (MS) with ion traps, operating in both positive and negative ESI mode and allowing for automated MS/MS fragmentation (Scheme 1). The separation is achieved by applying a smooth water–acetonitrile gradient on a reversed phase column. Retention times (R_t) of all detectors are corrected by comparison with external standards. Therefore, the actual retention times of eight reference compounds with different polarities are set to their initially defined values by multiplication with linear correction factors for the respective sections of the chromatograms.[15] This procedure leads to reproducible R_t values on different charges of a column type or on different machines, in our experience usually within ± 0.2 min. Only for compounds that are more lipophilic than the most lipophilic reference, retention times cannot be interpolated but are extrapolated instead, which may result in slightly higher deviations. After the correction, the peaks in all chromatograms are automatically assigned and integrated. For each peak, molecular weight values (AutoMW) are calculated by an algorithm with respect to the most common adducts or losses for both positive and negative ESI mode (*e.g.* $+H^+$ or $+Na^+$ in positive, $-H^+$ or $HCOO^-$ in negative ESI mode).[16]

The collated information of UV/Vis spectra, (HR)MS and MS/MS spectra, retention times (R_t) from DAD and/or ELSD chromatograms, and AutoMW is subsequently used for a single search in a reference library. This steadily growing database was set up on the basis of IMD's pure compound collection and now contains the information of approximately 15,000 unique NPs, among them the most prominent and abundant representatives of both plant and microbial bioactive metabolites, and a significant proportion of novel structures according to literature. An individual weighting of each search parameter is possible but in general not mandatory. However, it may be useful to reduce the weighting of UV/Vis data for peaks corresponding to minor components, due to concentration-dependence of the respective spectra. Or, if the regular MS spectrum seems to be superposed by a more easily ionizable minor component, one might want to neglect these data in order to obtain better results, or to focus on MS/MS data belonging to the component of interest. In any case, the result is presented as a hit list, with the single entries along with their structural formulae sorted by a quality factor calculated over all search parameters. Furthermore, the user can either apply the dereplication procedure on a specified peak or perform a batch dereplication, which automatically determines the best database hits for all peaks of a given chromatogram.

Another concept of working with the acquired data is the search for structural analogues. This is used to obtain information on the structural class of a specific compound if the exact reference is not included in the database. For this task, the retention times, MS spectra and AutoMW should be ignored, and the search strategy should focus on the UV spectrum, possibly with consideration of MS/MS data. This approach apparently relies on the quality of the UV spectra. Additionally, purpose-built algorithms have been proposed in literature.[17] However, if the analogue is not available as database reference, further NMR spectroscopic support and investigation is necessary to clearly and fully establish the molecular structure.

The described approach is exemplified by the secondary metabolite profiling of xylariaceaeous ascomycetes, being part of an extensive study on chemotaxonomy carried out in the past years in our group in collaboration with mycologists and natural products chemists from around the world. Despite that more than 50 novel metabolites have been identified from these previously neglected species during the past five years,

there are still numerous species that were revealed to contain unprecedented peaks by HPLC profiling. Compounds that appeared to be unique or characteristic of certain species or species groups were first recognized by HPLC-profiling, and thereafter the fruit bodies were collected in sufficient quantities for preparative work.[18] IMD also disposes of bioactivity data in up to over 100 assays for most samples of our extract and pure compound libraries, which have generally not been published yet. The example provided below is therefore merely intended to illustrate the dereplication process that is carried out as a routine procedure with all screening hits, preceding the bioactivity-guided isolation.

Fig. 1 shows the processed (collated, integrated, R_t corrected) chromatographic traces of a methanolic extract of *Hypoxylon fragiforme*. A subsequent batch dereplication performed on the combined spectroscopic information and calculated AutoMWs resulted in hit tables for each metabolite. For each single hit, a quality factor indicated the total similarity of actual and reference data (*e.g.* Fig. 2). By visual comparison of the data of both the respective peak and the database reference (Fig. 3), the identity of components **1**–**5** (Fig. 4) was confirmed as proposed by the quality factor. The presence of orsellinic acid (**1**) and four metabolites of the mitorubrin family (**2**–**5**) resembles well our previous observation that this type of metabolite is widely distributed among this species.[18] The lipophilic compounds **X** and **Y** did not match any database entry with sufficient quality and remain unidentified.

Fig. 5 shows the HPLC profile of a stromatal extract of *Annulohypoxylon urceolatum* YMJ90112618 from Taiwan,[19] which

	Quality	Comp. ID	Comp. Name	HR mass	MW	Retention Time
1	0,921432	NP_Structure	IMD 005474	440,110735	440,4020	14,10
2	0,520369	NP_Structure	IMD 011353	440,252253	440,5334	9,02
3	0,452816	NP_Structure	IMD 009769	440,131865	440,3986	9,62
4	0,428359	NP_Structure	IMD 000694	440,292660	440,6200	22,59
5	0,311295	NP_Structure	IMD 000977	440,292660	440,6200	23,10

Fig. 2. Hitlist for compound **4** from *Hypoxylon fragiforme*

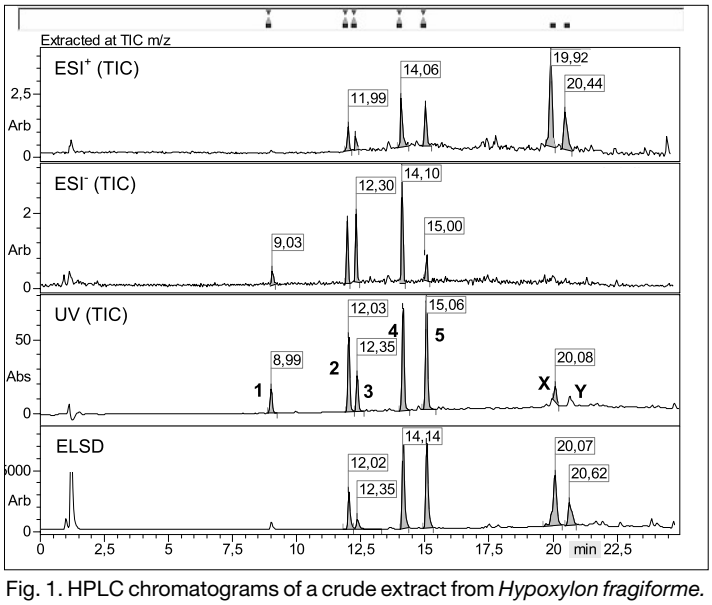


Fig. 1. HPLC chromatograms of a crude extract from *Hypoxylon fragiforme.* Compounds **1**–**5** were identified by database search, whereas no significant similarity was found for peaks **X** and **Y**. The upper result bar indicates the hit qualities (upper triangle >0.8, lower triangle >0.5, square >0.2).

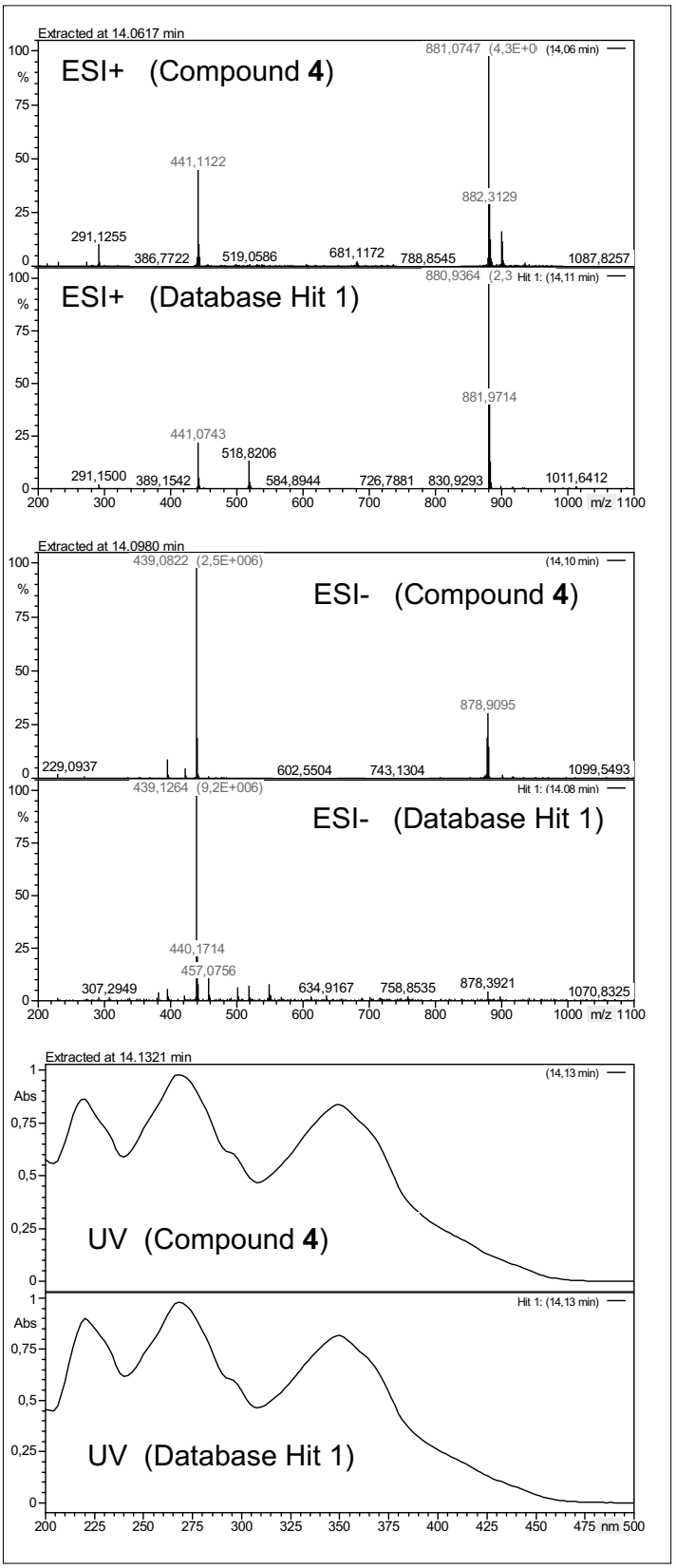


Fig. 3. Comparison of spectral data of compound **4** and the best database hit (quality 0.92)

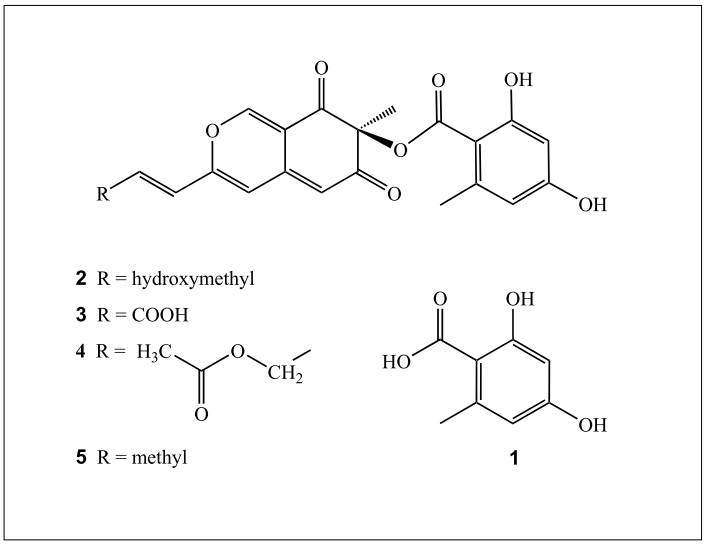


Fig. 4. Structural formulae of orsellinic acid (**1**), mitorubrinol (**2**), mitorubrinic acid (**3**), mitorubrinol acetate (**4**), and mitorubrin (**5**)

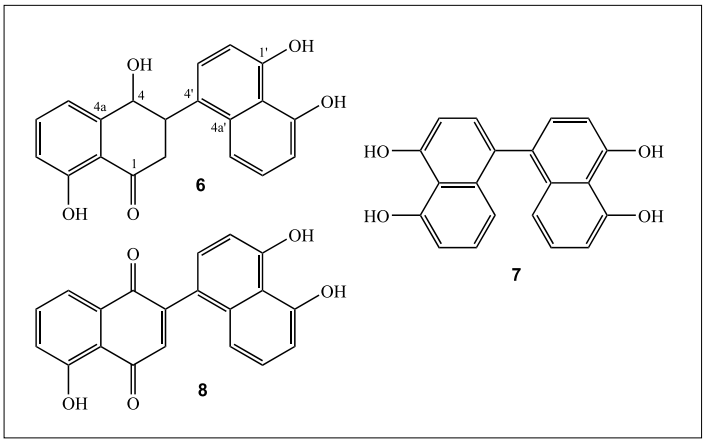


Fig. 6. Structural formulae of urceolone (**6**), BNT (**7**), and hypoxylone (**8**) from *A. urceolatum*

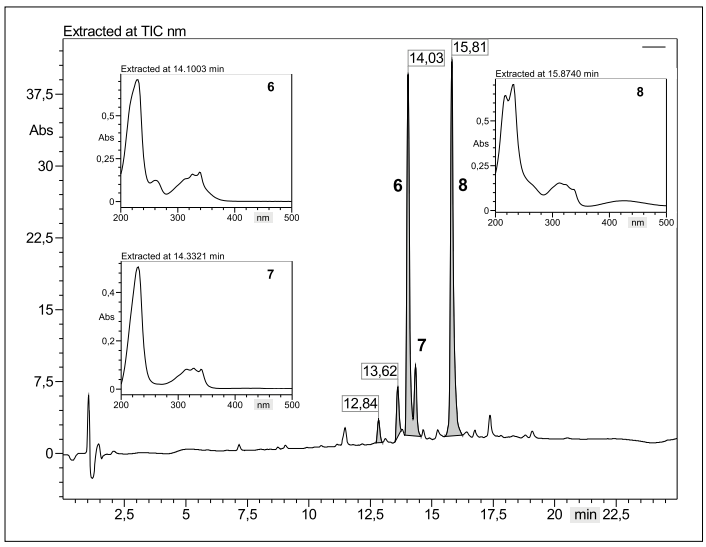


Fig. 5 UV/Vis chromatogram (TIC) of an extract from *A. urceolatum* and the UV spectra of compounds **6**, **7**, and **8**, that were identified as urceolone, BNT, and hypoxylone, respectively

Table 1. ESI-MS and NMR data (DMSO-d_6, 500 MHz) of urceolone (**6**)

pos.	δ_C [a]	δ_H, mult. (J_{HH} [Hz])
1	206.3	
2	38.4	2.62, dd (16.8, 12.8) 3.59, dd (16.8, 3.3)
3	39.5	4.22, br d (12.6)
4	68.7	4.79, s OH: 5.37, br s
4a	147.3	–
5	119.8	6.96, d (7.9)
6	137.2	7.59, t (7.9)
7	117.0	6.95, d (7.9)
8	162.9	OH: 12.37, s
8a	115.0	–
1’	153.3	OH: 10.97, s
2’	108.3	6.75, d (7.8)
3’	126.8	7.30, d (7.8)
4’	114.8	–
4a’	134.6	–
5’	114.5	7.45, d (8.0)
6’	127.1	7.32, t (7.9)
7’	108.6	6.78, d (7.9)
8’	153.2	OH: 11.01, s
8a’	114.8	–

ESI-MS; pos. ions *m/z* 319 $[M\text{-}H_2O+H]^+$, 337 $[M+H]^+$; neg ions *m/z* 335 $[M\text{-}H]^-$.

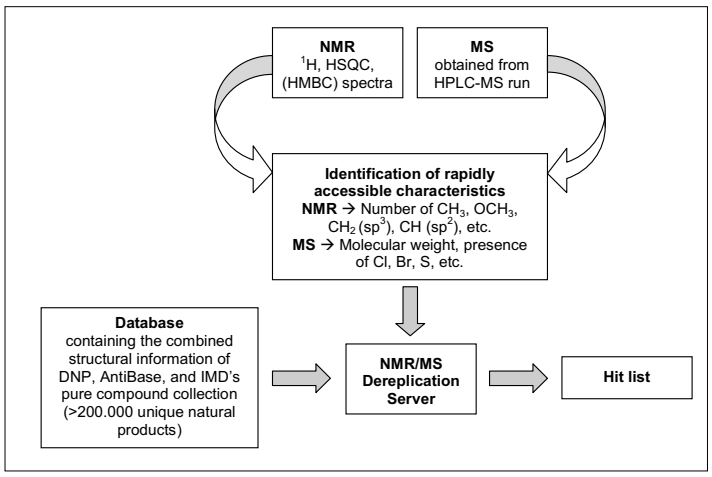
HRESI-MS; pos. ions *m/z* 319.0968 $[M\text{-}H_2O+H]^+$ (calcd for $C_{20}H_{15}O_4$, 319.0970).

[a] ^{13}C NMR shifts were obtained from the HSQC and HMBC correlation spectra.

had so far not been subjected to preparative work. From a comparison with the aforementioned spectral data on other Xylariaceae metabolites, it became evident that the extract contained, besides 1,1’,4,4’-tetrahydroxybinaphthalene (BNT, **7**), two unprecedented metabolites that are apparently specific for the species. These compounds (**6**, **8**) were isolated to purity by preparative HPLC. Whereas **8** proved identical to hypoxylone,[20] previously isolated from *Hypoxylon sclerophaeum*, **6** constitutes a new natural product for which we propose the trivial name urceolone (Fig. 6.). NMR and MS data are compiled in Table 1. Using the pure urceolone as standard, we have meanwhile established that the compound is not present in over 3000 extracts from various Xylariaceae, which demonstrates that our HPLC dereplication system is rather powerful for detection of unknown natural products. Chemotaxonomic significance of these findings will be explored elsewhere.

3. Advanced Library Supported Dereplication of Purified Natural Products by NMR and LCMS

As described above, the identification of metabolites by multidetector HPLC techniques is principally limited by the size of the reference database. Accordingly, if a specific substance was not found by the LCMS-based multidetector dereplication system, it could either be new or belong to the non-covered part of known NPs. Traditionally, structure elucidation is a time-consuming process. It usually requires a full set of 1D and 2D NMR data (^1H, ^{13}C, $^1H^1H$-COSY, $^1H,^{13}C$-HSQC, $^1H,^{13}C$-HMBC spectra), and an interpretable mass spectrum. Although some computer-assisted approaches have been proposed,[21] it still is a mainly hand-driven procedure that cannot be automated and demands the experience of a scientific expert. Thus, it is desirable to focus on those molecules which are new and avoid expedient work on already known



Scheme 2. Rapid identification of literature-known metabolites by NMR/MS dereplication

Table 2. Searchable fragments derived from NMR data

Type	Subtypes	Observed by	Main identification parameters
CH_3	Alkyl-CH_3, O-CH_3, N-CH_3	HSQC, (HQQC)	Distinct chemical shift, phase, intensity, (four quantum filtration)
CH_2	sp^2, sp^3	HSQC	Chemical shift and phase information
CH	sp^2, sp^3	HSQC	Combined chemical shift, intensity, and phase information
$R_2C=O$	–	HMBC	Chemical shift (δ_C >180)
CHO	–	HSQC, HMBC	Chemical shift (δ_C >190), attached proton
CO_2R	ester, carboxylic acid	HMBC	Chemical shift (160 < δ_C < 180), may be ambiguous
CON (amide)	–	HMBC	Chemical shift (160 < δ_C < 180), may be ambiguous
$OC(R_2)O$ (ketal)	–	HMBC, (HSQC)	Chemical shift, ambiguous unless other information available (*e.g.* anomeric proton)

structures. This implies the need to quickly elucidate whether or not a compound has been described in literature before, while requiring only a minimum amount of spectroscopic information and interpretation, and being independent from the physical availability of the authentic reference.

IMD's proprietary solution for the described challenge is a method we call NMR/MS dereplication (Scheme 2). As shown by Bradshaw and co-workers,[22] an exact count of methyl, methylene, and methine groups in combination with the molecular weight is usually sufficient to break down the number of literature-known structures that fit these criteria to less than ten. Our approach generalizes and extends this idea, leading in many cases to an unequivocal identification. It requires a mass spectrum (usually obtained from a LCMS run) and at least a 1D proton and a 2D HSQC NMR spectrum, from which the presence of characteristic groups like methyl, methoxy, methine, methylene (sp^2 and sp^3) or anomeric protons can easily be extracted (see Table 2 for a complete list), based on their characteristic chemical shifts and multiplicities. The ^{1}H,^{13}C-HSQC spectrum must be recorded with a pre- and post-evolution delay of ~$1/^2J_{CH}$ to allow sign discrimination comparable to a 1D DEPT spectrum.[23] Additional information on carbonyl groups like esters, amides, carbon acids and ketones can be obtained from an HMBC experiment and will help to discriminate between similar isoweight structures. For all identified moieties, their total number can either be counted or, in case of ambiguous data, described with the mathematical operators >, >=, <, <=. This is particularly useful for overlapping resonances or if methylene groups cannot be counted precisely by HSQC data, due to chemical shift degeneration. Fragments can also be excluded explicitly if they are obviously not present (*e.g.* no carbonyls if all resonances show δ_c <160). Apart from the molecular weight of a compound, the mass spectrum additionally bears information on the elemental composition: Chlorine and bromine can be identified by their characteristic isotopic patterns, the presence of nitrogen is indicated by an odd molecular weight, and sulphur sometimes gives a typical fragmentation. This evidence is additionally regarded in the NMR/MS dereplication process. If data from high resolution mass spectrometry (HRMS) are available, it can also be included.

The rapidly extractable information from NMR and MS data is collated in a client application (Fig. 7) and processed by a server-based dereplication engine. Its database consists of all literature-known natural products, based on the latest editions of Chapman & Halls 'Dictionary of Natural Products'[24] and AntiBase,[25] and combined with the structures of IMD's in-house NP collection. For each structure, searchable secondary data was extracted containing average and monoisotopic molecular weight, elemental composition, and the occurrence of fragments as listed in Table 2. This allows the scientist to search within more than 200,000 unique structures by just one click. The result is obtained as a hit list with structures and names of molecules that fit the given criteria. If the input was not sufficient for unambiguous identification, the user may specify additional fragments and repeat the procedure. Finally, the accurate verification of the proposed structure on the basis of present NMR data is a usually fast process. This step could even be semi-automated by using commercially available software, relying on comparison of actual NMR data and chemical shift prediction.[26] Alternatively, routine acquisition of COSY and HMBC spectra will accelerate and facilitate thorough structure verification.

An example demonstrates the utility of this approach. A compound produced by the fungus *Sarawakus britannicus*, strain CBS 253.62, was not identified in the HPLC-based first dereplication step. Its molecular weight was determined as 266 g/mol, and its UV spectrum showed maximum absorption at 220 and 325 nm. After isolation of the pure substance, a ^{1}H and a HSQC NMR spectrum were recorded (Fig. 8). The following fragments could be determined without in-depth interpretation of the spectra: a methoxy group as the only methyl moiety present, two methylene groups (none of which is olefinic), a sp^3-hybridized methine group, and three olefinic or aromatic protons. These fragments and the molecular weight were used as query in the NMR/MS dereplication and revealed a single hit (Fig. 7). The structure resembled well the expected oxygenation of both the aliphatic methine and one of the methylene groups. Verification of the connectivities by COSY and HMBC data proved the identity with diaportinol, which has been reported as metabolite of *Penicillium nalgiovense*.[27] Another example, showing the feasibility of the NMR/MS dereplication concept as well for more complex structures, is given with the identification of cytochalasin H, which was isolated from a strain of *Hypoxylon fragiforme*. The HSQC spectrum again provided the information on the number of methyl, methylene, and methine groups (Fig. 9), among these a sp^2-hybridized methylene from an exocyclic double bond. The odd molecular weight (493 g/mol) indicated the presence of at least one

nitrogen atom. Other structural features like methoxy groups or halogens could be excluded by NMR and MS data, respectively. A dereplication search with the combined information returned cytochalasin H as the only known compound fitting the structural requirements. Verification was achieved by analyzing the connectivity information of the 2D NMR spectra, and comparison with literature data.[28]

4. Conclusions, Discussion and Perspectives

In the present paper we described a feasible and straightforward strategy for accelerated dereplication of natural products. A two-level approach is applied to attain a maximum of support by reference libraries at each stage while requiring a minimum of time for data collection and interpretation. The result is maximum compatibility with the narrow timelines of modern HTS-based discovery projects. The HPLC multidetector system as the first level can be used for crude extracts as well as for pure compounds. With the availability of hyphenated techniques like MS/MS, the continuous comparison with external standards and the ability to search all data simultaneously, it is ensured that the main limitation is only given by the size of the database. Hence, the success in this step inevitably relies on the number of available reference compounds. Our methodology based on high-resolution mass data from an HPLC-coupled ESI-TOF spectrometer provides an advanced solution here, as it allows natural product databases to be searched for HRMS data or molecular formulae, respectively. However, additional information on taxonomy and bioactivity as well as sophisticated data interpretation is inevitably needed for the determination of a compound without reference HPLC data. One problem of the HPLC approach is that for efficiency reasons it is typically limited to a single chromatographic condition. As no condition is equally suitable for all different types of biomolecular structures, this inherently means that for some examples the achieved separation, for instance of two closely related regioisomers, may not be good enough to unequivocally differentiate between two or more structures having the same mass and same UV absorption. If neither MS/MS nor phylogenetic data provide help, this will imply the need for a further NMR spectroscopic investigation.

The obstacle of limited availability of reference compounds can be overcome in the second-level process, the NMR/MS dereplication. It is applied to the purified samples and is independent from authentic references. Its goal is the accelerated recognition of any literature-known structure. Modern NMR spectrometers allow for a rapid acquisition of the needed experiments from (sub)-milligram amounts of a compound. As only basic interpretation of the spectra is necessary, even low-resolution data or samples not pure enough for detailed analysis will usually be sufficient. In these cases, acquisition of a ^{1}D HQQC spectrum[29] can be useful for unambiguous detection and count of methyl groups. This additional spectrum is recorded in only a few minutes and does not seriously restrict spectrometer time.

It should be mentioned that both the HPLC-based and the NMR-based dereplication methods are principally blind to questions of absolute stereochemistry. In case of the NMR/MS dereplication, the approach primarily focuses on determination of the composition, with some additional respect to conformation. For the unequivocal differentiation of epimers, one has ei-

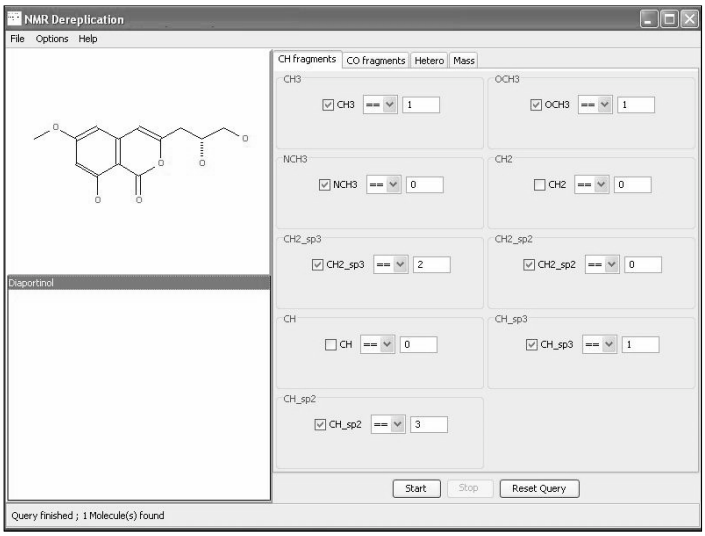


Fig. 7. NMR/MS dereplication of a fungal metabolite using basic NMR and MS data

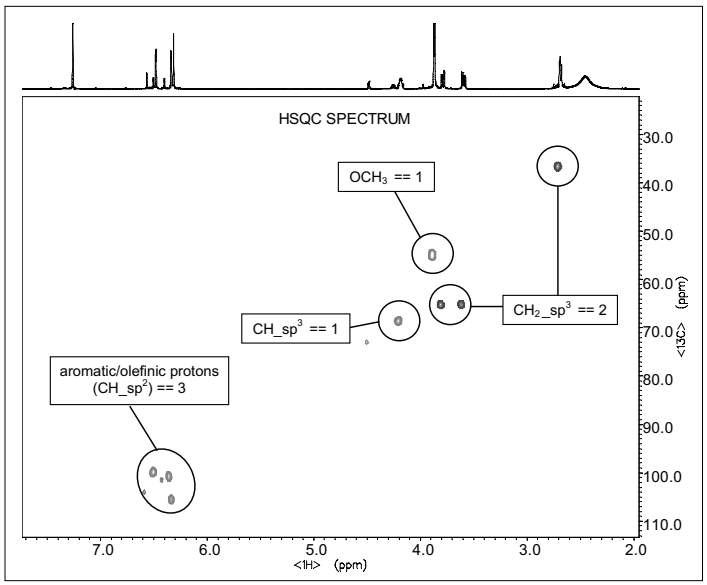


Fig. 8. HSQC NMR spectrum of a fungal metabolite, identified as diaportinol by NMR/MS dereplication

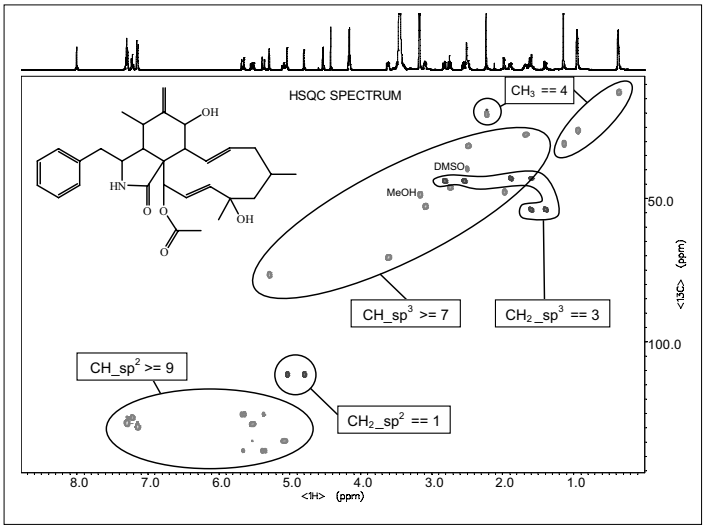


Fig. 9. Interpreted HSQC NMR spectrum and structural formula of cytochalasin H.

ther to compare with literature NMR data or to use additional spectroscopic information like coupling constants from ^{1}H NMR spectra, distance information from NOE measurements, CD spectra, or optical rotation values. The two latter techniques additionally would give access to the determination of enantiomers.

From a fundamental research point of view, the application of LC-hyphenated techniques involving NMR, OR, and CD to large screening libraries comprising several ten thousands of samples may look promising.[30] However, it should always be considered that dereplication is intended to save money, to avoid wasting resources, and to compete with synthetic lead-finding processes. Implementation of such cost- and time-intensive processes in the NP-related workflow for lead discovery therefore may have contributed to the reservations against NPs in the pharmaceutical industry, some of which tended to rely solely on combinatorial chemistry approaches. This, on the other hand, has led to empty pipelines in many companies who abandoned natural products research. In contrast, our own approach is designed to guarantee a maximum degree of innovation, using instruments and human capital in a cost-efficient way.

As a future outlook, a combination of the NMR/MS dereplication strategy with HPLC-SPE-NMR techniques seems capable. It could overcome the problem of how to interpret these NMR data. Spectral databases are difficult to establish, due to variability of spectral results and resolution on the actual conditions, the need to suppress large solvent peaks, and the unavailability of literature NMR data in the mixed solvent systems used for HPLC. Moreover, time-consuming experiments like HMBC usually cannot be obtained. Instead, ^{1}H, HSQC, and possibly 1D HQQC spectra may provide sufficient information to extract characteristic fragments as described before. This approach could further be enhanced by adding proton multiplicities to the NMR/MS dereplication database or the use of chemical shifts obtained from public databases or calculations.

In summary, we showed how sophisticated dereplication strategies can accelerate the identification of natural products, making them an important and feasible tool for pharmaceutical and agrochemical discovery projects. Equally, scientific investigations like chemotaxonomy or biological profiling gain profit from the advanced techniques.

Received: April 3, 2007

[1] a) M. S. Butler, *J. Nat. Prod.* **2004**, *67*, 2141; b) D. J. Newman, G. M. Cragg, K. M. Snader, *Nat. Prod. Rep*. **2000**, *17*, 215; c) J. Clardy, C. Walsh, *Nature* **2004**, *432*, 829.

[2] a) S. Saxena, A. K. Pandey, *Appl. Microbiol. Biotechnol.* **2001**, *55*, 395; b) 'Crop Protection Agents from Nature', Ed. L.G. Copping, The Royal Society of Chemistry, Cambridge, UK, **1996**; c) J. P. Pachlatko, *Chimia* **1998**, *52*, 29.

[3] a) D. T. Hung, T. F. Jamison, S. L. Schreiber, *Chem. Biol.* **1996**, *3*, 623; b) K. Stromgaard, *Chem. Rec.* **2005**, *5*, 229.

[4] a) A. Harvey, *Drug Discovery Today* **2000**, *5*, 294; b) J. Berdy, *J. Antibiot.* **2005**, *58*, 1.

[5] a) T. Henkel, R. M. Brunne, H. Müller, F. Reichel, *Angew. Chem.* **1999**, *111*, 688; *Angew. Chem., Int. Ed.* **1999**, *38*, 643; b) J. Larsson, J. Gottfries, L. Bohlin, A. Backlund, *J. Nat. Prod*. **2005**, *68*, 985.

[6] V. Knight, J.-J. Sanglier, D. DiTullio, S. Braccili, P. Bonner, J. Waters, D. Hughes, L. Zhang, *Appl. Microbiol. Biotechnol*. **2004**, *62*, 446.

[7] a) M. L. Sogin, H. G. Morrison, J. A. Huber, D. M. Welch, S. M. Huse, P. R. Neal, J. M. Arrieta, G. J. Herndl, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 12115; b) N. R. Pace, *Science* **1997**, *276*, 734; c) T. P. Curtis, W. T. Sloan, J. W. Scannell, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 10494.

[8] S. D. Bentley, K. F. Chater, A.-M. Cerdeño-Tárraga, G. L. Challis, N. R. Thomson, K. D. James, D. E. Harris, M. A. Quail, H. Kieser, D. Harper, A. Bateman, S. Brown, G. Chandra, C. W. Chen, M. Collins, A. Cronin, A. Fraser, A. Goble, J. Hidalgo, T. Hornsby, S. Howarth, C.-H. Huang, T. Kieser, L. Larke, L. Murphy, K. Oliver, S. O'Neil, E. Rabbinowitsch, M.-A. Rajandream, K. Rutherford, S. Rutter, K. Seeger, D. Saunders, S. Sharp, R. Squares, S. Squares, K. Taylor, T. Warren, A. Wietzorrek, J. Woodward, B. G. Barrell, J. Parkhill, D. A. Hopwood, *Nature* **2002**, *417*, 141.

[9] H. B. Bode, B. Bethe, R. Höfs, A. Zeeck, *ChemBioChem* **2002**, *3*, 619.

[10] J. Bok, D. Hoffmeister, L. Maggio-Hall, R. Murillo, J. Glasner, N. Keller, *Chem. Biol.* **2006**, *13*, 31.

[11] a) G. A. Cordell, Y. G. Shin, *Pure Appl. Chem.* **1999**, *71*, 1089; b) M. A. Strege, *J. Chromatogr. B* **1999**, *725*, 67; c) K. F. Nielsen, J. Smedsgaard, *J. Chromatogr. A* **2003**, *1002*, 111.

[12] a) A. Fredenhagen, C. Derrien, E. Gassmann, *J. Nat. Prod.* **2005**, *68*, 385; b) Y. Konishi, T. Kiyota, C. Draghici, J. Gao, F. Yeboah, S. Acoca, S. Jarussophon, E. Purisima, *Anal. Chem*. **2007**, *79*, 1187.

[13] a) S. C. Bobzina, S. Yangb, T. P. Kastena, *J. Chromatogr. B* **2000**, *748*, 259; b) V. Exarchou, M. Krucker, T. A. van Beek, J. Vervoort, I. P. Gerothanassis, K. Albert, *Magn. Reson. Chem*. **2005**, *43*, 681; c) J. W. Jaroszewski, *Planta Med.* **2005**, *71*, 691.

[14] M. Stadler, V. Hellwig, in 'Handbook of Industrial Mycology', Eds. Z. An, Marcel Dekker, New York, **2004**, p. 269.

[15] The retention times of a section are corrected using the following formula:

$$R_{t,corr} = R_{t,R1} + \frac{\Delta R_t}{\Delta R_{t,act}} * \left(R_{t,act} - R_{t,R1\_act}\right)$$

wherein $R_{t,R1}$ and $R_{t,R2}$ are the reference retention times of the two standards delimiting the section, and $R_{t,R1\_act}$ and $R_{t,R2\_act}$ are their actual values. $\Delta R_t = (R_{t,R2} - R_{t,R1})$, $\Delta R_{t,act} = (R_{t,R2\_act} - R_{t,R1\_act})$. $R_{t,act}$ is the obtained retention time of a peak, and $R_{t,corr}$ is its corrected value. This takes up a concept described in: S. P. Elliott, K. A. Hale, *J. Chromatogr. B* **1997**, *694*, 99.

[16] H. Tong, D. Bell, K. Tabei, M. M. Siegel, *J. Am. Soc. Mass Spectrom.* **1999**, *10*, 1174.

[17] T. Ostenfeld Larsen, B. O. Petersen, J. O. Duus, D. Sørensen, J. C. Frisvad, M. E. Hansen, *J. Nat. Prod.* **2005**, *68*, 871.

[18] M. Stadler, J. Fournier, *Rev. Iberoam. Micol.* **2006**, *23*, 160 (available at *http://www.reviberoammicol.com/2006-23/160170.pdf*); b) M. Stadler, J. Fournier, D.N. Quang, A.Y. Akulov, *Nat. Prod. Commun*. **2007**, *2*, 287; c) M. Stadler, V. Hellwig, *Recent Res. Devel. Phytochem.* **2005**, *9*, 41.

[19] D. N. Quang, T. Hashimoto, Y. Nomura, H. Wollweber, V. Hellwig, J. Fournier, M. Stadler, Y. Asakawa, *Phytochemistry* **2005**, *66*, 797.

[20] B. Bodo, R. G. Thi, D. Davoust, H. Jacquemin, *Phytochemistry* **1983**, *22*, 2579.

[21] a) C. Steinbeck, *Nat. Prod. Rep.* **2004**, *4*, 512; b) M. E. Elyashberg, K. A. Blinov, A. J. Williams, E. R. Martirosian, S. G. Molodtsov, *J. Nat. Prod.* **2002**, *65*, 693.

[22] J. Bradshaw, D. Butina, A. J. Dunn, R. H. Green, M. Hajek, M. M. Jones, J. C. Lindon, P. J. Sidebottom, *J. Nat. Prod*. **2001**, *64*, 1541.

[23] D. Neuhaus, J. Keeler, R. Freeman, *J. Mag. Res*. **1985**, *61*, 553.

[24] 'Dictionary of Natural Products', on CD-ROM, Version 15:1, Chapman & Hall/CRC, Boca Raton, USA, **2006**.

[25] H. Laatsch, 'AntiBase 2006', a database for the identification of natural products, Chemical Concepts, Weinheim, Germany, **2006**.

[26] a) S. S. Golotvin, E. Vodopianov, B. A. Lefebvre, A. J. Williams, T. D. Spitzer, Magn. Reson. Chem. **2006**, *44*, 524; b) L. Griffiths, R. Horton, *Magn. Reson. Chem*. **2006**, *44*, 139.

[27] T. Ostenfeld Larsen, J. Breinholt, *J. Nat. Prod.* **1999**, *62*, 1182.

[28] M. Stadler, D. N. Quang, A. Tomita, T. Hashimoto, Y. Asakawa, *Mycol. Res.* **2006**, *110*, 811.

[29] M. Liu, R. D. Farrant, J. K. Nicholson, J. C. Lindon, *J. Magn. Reson. Series B* **1995**, *106*, 270.

[30] G. Bringmann, K. Messer, M. Wohlfarth, J. Kraus, K. Dumbuya, M. Rückert, *Anal. Chem.* **1999**, *71*, 2678.