Published in final edited form as:

Phytochemistry. 2015 May ; 113: 87–95. doi:10.1016/j.phytochem.2014.11.020.

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

Justin T. Fishedick¹, Sean R. Johnson, Raymond E.B. Ketchum², Rodney B. Croteau, and B. Markus Lange*

Institute of Biological Chemistry and M.J. Murdock Metabolomics Laboratory, Washington State University, Pullman, WA 99164-6340, USA

Abstract

Development and testing of Spektraris-NMR an online spectral resource, is reported for the NMR-based structural identification of plant natural products (PNPs). Spektraris-NMR allows users to search with multiple spectra at once and returns a table with a list of hits arranged according to the goodness of fit between query data and database entries. For each hit, a link to a tabulated alignment of ¹H-NMR and ¹³C-NMR spectroscopic peaks (query versus database entry) is provided. Furthermore, full spectroscopic records and experimental meta information about each database entry can be accessed online. To test the utility of Spektraris-NMR for PNP identification, the database was populated with NMR data (total of 466 spectra) for ~250 taxanes, which are structurally complex diterpenoids (including the anticancer drug taxol) commonly found in the genus *Taxus*. NMR data generated used was then generated with metabolites purified from *Taxus* cell suspension cultures to search Spektraris-NMR, and were able to identify eight taxanes with high confidence. A ninth isolated metabolite could be assigned, based on spectral searches, to a taxane skeletal class, but no high confidence hit was produced. Using various spectroscopic methods, this metabolite was characterized as the taxane 2-deacetyl baccatin IV, a novel taxane. These results indicate that Spektraris-NMR is a valuable resource for rapid and reliable identification of known metabolites and has the potential to contribute to de-replication efforts in the search for novel PNPs.

Keywords

database; metabolomics; natural product; nuclear magnetic resonance; taxol; *Taxus*

© 2014 Elsevier Ltd. All rights reserved.

*Corresponding author. Tel.: xx1-509-335-3794; fax: xx1-509-335-7643. lange-m@wsu.edu (B.M. Lange).

¹Present address: Excelsior Analytical, 30099 Ahern Ave., Union City, CA 94587.

²Present address: Diana Plant Sciences, 18167 S.W. Boones Ferry Rd., Portland, OR 97224, USA

Contribution on the occasion of Professor Vincenzo de Luca's 60th birthday

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

1. Introduction

Plants are a remarkable source of chemically diverse small molecules that are often referred to as secondary metabolites, specialized metabolites or plant natural products (PNPs). Isolated, biologically active, PNPs have a long history of use in the food, flavor, fragrance and pharmaceutical industries (Fortineau, 2004; Li and Vederas, 2009; Jones et al., 2012). The rapid and reliable identification of PNPs is of critical importance for the chemical screening of complex plant extracts. Hyphenated techniques, which combine the separation of metabolites by high performance liquid chromatography (HPLC) with online detection by mass spectrometry (MS), have emerged as some of the most versatile analytical approaches. Various searchable online database resources have been developed for HPLC-MS and MS/MS fragmentation data obtained with PNPs (Smith et al., 2005; Horai et al., 2010; Sawada et al., 2012; Cuthbertson et al., 2013; Fukushima and Kusano, 2013). Nuclear magnetic resonance (NMR) spectroscopy is another essential technique, complementary to HPLC-MS, for PNP structure elucidation. Several online NMR spectroscopic databases and their prediction tools are now available, including SDBS (Yamamoto et al., 1988), NMRShiftDB (Steinbeck et al., 2003; Steinbeck and Kuhn, 2004), HMDB (Wishart et al., 2007), nmrdb.org (Banfi and Patiny, 2008), MMCD (Cui et al., 2008), and BML-NMR (Ludwig et al., 2012). However, these web-based NMR resources focus on synthetic organic compounds or human metabolites, and coverage of PNPs, in particular those of low abundance in the source plant(s), is very limited. The majority of PNPs are not available commercially, which is, at least in part, due to the fact that small amounts of a particular metabolite may have been isolated decades ago and authentic samples are not available any more. A re-isolation or re-synthesis of a larger number of PNPs is not practical. The Spektraris toolbox (<http://langelabtools.wsu.edu/Spektraris>), which until now focused on HPLC-MS data sets for PNPs (based on the accurate mass-time tag approach; Cuthbertson et al., 2013), was thus expanded to include a searchable database for NMR spectroscopic data that were extracted from the literature. As a proof-of-concept, the utility of the Spektraris-NMR resource was tested to identify various taxanes, which belong to a structurally complex family of diterpenoid metabolites commonly found across the genus *Taxus* (yew).

Over 500 naturally occurring taxanes have been described in the literature (Wang et al., 2011), with paclitaxel (taxol®) (**1**) (Fig. 1), a blockbuster anticancer drug (Cragg and Newman, 2013), as a signature member of this metabolite family. Several approaches have been explored to avoid re-identifying previously characterized taxanes (de-replication). HPLC operating in tandem with ¹H-NMR spectroscopy has been demonstrated to be useful for this purpose (Schneider et al., 2003; Eldridge et al., 2002). A computerized library of ¹³C-NMR spectroscopy data for taxanes in European yew (*Taxus baccata* L.) extracts was also reported (Duquesnoy et al., 2009). Unfortunately, the data for these studies are neither accessible online nor downloadable for offline searching. As a proof-of-concept study, 9 taxanes were isolated from cell suspension cultures of *Taxus × media* cv. Hicksii, the accurate mass and NMR spectra of which were then compared against our Spektraris database (which contains NMR spectroscopic data for about 250 taxanes). Using this approach, eight known taxanes were identified with high confidence (**1** (taxol), **2** (baccatin III), **3** (*N*-acetyl-*N*-debenzoyltaxol), **4** (13-acetyl-9-dihydrobaccatin III), **5** (baccatin IV), **6**

(taxol D), **7** (taxol B), **8** (taxol C)) (Fig. 1), and one additional taxane to a structural class with a 6/8/6 ring system. The unknown taxane was characterized, based on additional ^1H - ^1H COSY, DEPT, HSQC, HMBC, and NOESY experiments, as a novel metabolite, 2-deacetylbaicatin IV (**9**).

2. Results and discussion

2.1. Spektraris-NMR database design

NMR spectroscopic data for 247 taxanes, mostly containing the relatively abundant 6/8/6 ring system, were extracted (mostly manually) from the published literature and entered into a Microsoft Excel file. ^1H and ^{13}C -NMR spectroscopic data were available for 217 metabolites, while only ^1H -NMR or ^{13}C -NMR data were accessible for 28 and 2 metabolites, respectively (total of 466 spectra). The data collected from published ^1H -NMR data (as searchable parameters in Spektraris) were chemical shifts (δ in ppm; acquired in CDCl_3 if available), peak types based on splitting caused by spin-spin coupling (s, d, dd, etc.), and coupling constants (J in Hz) (Fig. 2A). From published ^{13}C -NMR spectroscopic data, only chemical shifts were collected as searchable parameters. Focus was placed on NMR data acquired in CDCl_3 as a solvent and, to ensure consistency, suggest its use for user-acquired spectra to be integrated into Spektraris-NMR. The molecular formula, monoisotopic mass, name (and synonyms), and CAS registry number were either taken from the original paper (when available) or obtained by searching SciFinder[®]. Searches against other online chemical databases, such as ChemSpider or PubChem, were mostly unsuccessful as they contained only very few taxanes of interest. NMR spectroscopic and metabolite meta data were then loaded into an SQL relational database (Fig. 2A).

The search for spectral data in the published literature turned out to be a considerable challenge. Not all original papers were written in English and, in some cases, spectroscopic data included poorly resolved peaks because of the acquisition with older NMR instruments. To support attempts aimed at digitizing spectroscopic data, it would have been desirable if NMR data had been made available in a common, machine-readable format, as suggested by community initiatives (Rubtsov et al., 2007). Consistent with a free data exchange policy, the contents of the Spektraris-NMR database can be accessed at <http://langelabtools.wsu.edu/nmr/> without user registration. Spectra acquired locally are made available as the Varian binary (fid) file, JCAMP-DX (Davies and Lampen, 1993), and in the emerging standard format nmrML (<http://nmrml.org/>). To further facilitate data dissemination, the spectral analyses has also been submitted for integration into the NMRShiftDB resource (Steinbeck et al., 2003; Steinbeck and Kuhn, 2004). In addition to searching spectra in Spektraris, users can also browse a list of all compounds in the database, or download the source code and data files (incl. spectral details) used to build the database and webpage. New compounds can be submitted to the database by filling out an intuitive web-based form.

Distinctive features of Spektraris-NMR are tools that allow to combine (1) searching with multiple spectra at once, (2) ranking hits according to goodness of fit between query data and database entries, and (3) tabulating matches between query and database peaks. These combined capabilities have not yet been integrated into other databases. SDBS (Yamamoto

et al., 1988), for example, allows multiple spectra for the same compound to be searched together, but sorting according to goodness of match to the input is not enabled. The strategy of searching combinations of spectroscopic data and sorting hits by combined similarity was first proposed by Clerc and Erni (1973), and has been implemented in various tools, including SPECTRA (Masui and Yoshida, 1996) and MMCD (Cui et al., 2008). NMRShiftDB (Steinbeck et al., 2003; Steinbeck and Kuhn, 2004) sorts hits by goodness of match, and displays a peak matching table, but only allows searches on one kind of spectrum at a time. It is also important to note that Spektraris-NMR contains hundreds of PNP spectra not included in any other database.

In addition to the online tools mentioned above, several software packages for database searching are available through commercial vendors (e.g., ACD/Labs NMR Database, Bruker Amix, and Chenomx NMR Suite). The former uses an approach similar to ours (searching with extracted δ and J values), while the two latter packages allow direct searches with spectroscopic data. A direct comparison with Spektraris-NMR is not possible because these commercial databases contain just a handful of taxane spectra or none at all. This illustrates a common problem: the original NMR data files for tens of thousands of PNPs not sold by commercial laboratory supply companies are unavailable. The extraction of NMR chemical shifts and coupling constants from the older literature (as we have done for Spektraris-NMR), albeit tedious, is therefore the only practical solution to develop searchable databases.

2.2 Spektraris-NMR database searching

As a prerequisite for Spektraris-NMR searches, NMR spectroscopic data need to be arranged by chemical shifts (δ in ppm), peak types (splitting pattern) and/or coupling constants (J in Hz) (only chemical shifts for ^{13}C -NMR data), and can be pasted directly from Microsoft Excel (tab- or pipe-delimited format) (Fig. 2A, B). Tolerances can be included for the chemical shifts (the default values of 0.05 and 0.3 ppm for ^1H - and ^{13}C -NMR spectroscopy, respectively, generally worked well in the tests with taxane spectra). If known, the monoisotopic mass and its tolerance can be used as an independent parameter in the search (Fig. 2C). Spektraris will return search results as a list of hits with metabolite name, molecular formula, monoisotopic mass, scores for spectroscopic comparisons (^1H , ^{13}C and MS) and a depiction of the structure. The right-hand column of the results table contains peak alignment tabs, which link to side-by-side comparisons of ^1H (under “Proton”) or ^{13}C (under “Carbon”) query NMR spectroscopic data against the database records. The list can be sorted by the ^1H -NMR, ^{13}C -NMR, monoisotopic mass or average similarity score. When the metabolite name in the results list is clicked, the full database record is presented, including details of how the original NMR spectroscopic data were acquired, further information about the metabolite, a list of relevant references, a table with full peak annotations and data interpretation notes (Fig. 3D). To obtain independent structural evidence, users can also use HPLC-MS data obtained with high-resolution instruments to search the Spektraris-AMT tag library for natural products (Cuthbertson et al., 2013) or employ MS/MS spectra to search against the MassBank database (Horai et al., 2010) (Fig. 3B).

2.3 Isolation of taxane diterpenoids and structure elucidation based on Spektraris searches

Cells from *Taxus × mediacv. Hicksii* suspension cultures were harvested and extracted with CH₂Cl₂. After drying of the extract *in vacuo*, the residue was suspended in 50 % aqueous acetonitrile and injected onto a reversed phase HPLC column operated isocratically with acetonitrile/water (42:58; v/v). Fractions were collected and the injection/fractionation procedure repeated until the entire raw extract had been processed. Each fraction was evaluated for taxane content (based on signature mass spectral fragmentation patterns) and purity by analyzing an aliquot using rapid resolution HPLC-QTOF-MS. Where appropriate, fractions were combined. If necessary, fractions were subjected to a second HPLC-based fractionation, and these new fractions also analyzed by rapid resolution HPLC-QTOF-MS. For metabolite identification, relative retention times (referenced to anthracene 9-carboxylic acid as an internal standard) and m/z values (determined at high mass accuracy) were compared with those of a taxane authentic standard library (AMT tag approach analogous to Cuthbertson et al. (2013) but with a different chromatographic method). In ESI negative polarity mode, taxanes often formed formic acid adducts [M-H + C₁H₂O₂]⁻ as the base peak, a chloride adduct [M+Cl]⁻, and a smaller [M-H]⁻ ion (Supplementary Table 1). In APCI positive polarity mode, taxane spectra were characterized by a substantial [M+H]⁺ ion, typically accompanied by sodium [M+Na]⁺ and potassium [M+K]⁺ adducts. Various fragment ions were observed in APCI(pos) (Supplementary Table 1) and used as an independent criterion for metabolite identification. Based on AMT tag and MS/MS library comparisons, four metabolites could be identified with high confidence as taxane diterpenoids (**1**, paclitaxel or taxol ([M+H]⁺ = 854.3374, [M-H]⁻ = 852.3221; C₄₇H₅₁NO₁₄); **2**, baccatin III ([M+H]⁺ = 587.2481, [M-H]⁻ = 585.2341; C₃₁H₃₈O₁₁); **7**, taxol B (cephalomannine) ([M+H]⁺ = 832.3539, [M-H]⁻ = 830.3393; C₄₅H₅₃NO₁₄); and **8**, taxol C ([M+H]⁺ = 848.3852, [M-H]⁻ = 846.3706; C₄₆H₅₇NO₁₄)) (Fig. 4). However, authentic standards corresponding to the other metabolites isolated from *Taxus × media* cell cultures were not in our possession. NMR spectroscopic data were therefore acquired for these metabolites and to evaluate the utility of Spektraris-NMR for structural elucidation.

Spektraris searches with ¹H-NMR data and the molecular formulas calculated based on HPLC-QTOF-MS data resulted in one match for metabolites **3**, **4**, and **6**, which were thus assigned as *N*-acetyl-*N*-debenzoyltaxol ([M+H]⁺ = 792.3214, [M-H]⁻ = 790.3057; C₄₂H₄₉NO₁₄), 13-acetyl-9-dihydrobaccatin III ([M+H]⁺ = 631.2770, [M-H]⁻ = 629.2589; C₃₃H₄₂O₁₂), and taxol D ([M+H]⁺ = 820.3541, [M-H]⁻ = 818.3375; C₄₄H₅₃NO₁₄), respectively (Gunawardana et al., 1992; Appendino et al., 1994; Zhang et al., 2000). The same search for compounds **1** and **8** yielded three and two matches, respectively. However, the top hits scored at close to 100 % identity, which allowed the confirmation of **1** and **8** as taxol and taxol C, respectively, thereby confirming HPLC-QTOF-MS assignments (Chmurny et al., 1992; Barboni et al., 1994). Several hits were also returned in searches with ¹H-NMR spectra and the molecular formula data of metabolites **2**, **5**, and **7**. When the ¹³C-NMR spectroscopic data were included in the search, the top score unambiguously assigned the structures to baccatin III, baccatin IV ([M+H]⁺ = 653.2824, [M-H]⁻ = 651.2719; C₃₂H₄₄O₁₄) and taxol B, respectively, thereby confirming HPLC-QTOF-MS assignments of **2** and **7** (Rojas et al., 1983; Sénilh et al., 1984; Chmurny et al., 1992).

A search with ^1H -NMR and molecular formula data for $\mathbf{9}([\text{M}+\text{H}]^+ = 611.2687, [\text{M}-\text{H}]^- = 609.2528; \text{C}_{30}\text{H}_{42}\text{O}_{13})$ yielded six hits (taxumairol D, $1\beta,7\beta$ -dihydroxy- $4\beta,20$ -epoxy- $2\alpha,5\alpha,9\alpha,10\beta,13\alpha$ -pentaacetoxy-tax- 11 -ene, 7β -deacetoxy- $1\beta,7\alpha$ -dihydroxybaccatin I, 4-deacetyl baccatin IV, $1\beta,9\alpha$ -dihydroxy- $4\beta,20$ -epoxy- $2\alpha,5\alpha,7\beta,10\beta,13\alpha$ -pentaacetoxy-tax- 11 -ene, and taxuspine V). Five of these metabolites contain a C 4 ,C 20 -epoxide ring, while 4-deacetyl baccatin IV (**10**) has an oxetane ring in the same position. Proton signals at 4.63 ppm (1H, dd, $J = 8.9, 0.9$ Hz, H-20) and 4.58 ppm (1H, d, $J = 8.9$ Hz, H-20) suggested that **9** may also contain an oxetane ring. Based on the molecular formula deduced from the HPLC-QTOF-MS data, **9** likely contains five acetate groups. However, the low number of direct matches between proton chemical shifts of **9** and **10** (the latter of which has acetate groups on C-2, C-7, C-9, C-10 and C-13) was suggestive of a different acetylation pattern. When ^{13}C -NMR spectroscopic data for **9** were included in the Spektraris search, baccatin IV (**5**), which contains six acetate groups (on C-2, C-4, C-7, C-9, C-10, and C-13 of the taxane core), was returned as the top hit. Compound **9** was further characterized by $^1\text{H}^1\text{H}$ COSY, DEPT, HSQC, HMBC, and NOESY experiments (Table 1). The downfield shift of the H-2 signal of **9** (4.11 ppm; dd, $J = 5.5, 5.4$ Hz), when compared to **10** (2.69 ppm; d, $J = 5.5$ Hz), was indicative of the presence of a hydroxyl group at C-2. Further evidence for this assignment was provided by COSY correlations between H-2 of **9** and overlapping protons at 2.84 ppm (1-H, br d, $J \approx 6.0$ Hz; 1-H, d, $J \approx 5.2$). A correlation in HSQC between δC -3 47.23 (CH) and δH 2.84 (1-H, d, $J \approx 5.2$) confirmed the position of this proton at C-3, while the lack of correlation between C-3 and δH 2.84 (1-H, br d, $J \approx 6.0$ Hz) indicated that this proton belongs to the OH group at C-2. The remaining spin systems between H-9/10, H-5/2H-6/7, and H-13/2H-14 were confirmed by COSY and HMBC. Carbon signals were assigned based on a combination of DEPT, HSQC, and HMBC experiments (Table 1). The orientation of H-6 α was confirmed by NOESY correlations with H-5. Another NOESY correlation between H-6 α and H-7 confirmed an α -orientation for H-7. The β -orientation of H-9 was confirmed by NOESY correlation with the C-19 methyl group, while the α -orientation of H-10 was confirmed by NOESY correlations with H-7 and the C-16 methyl group. H-2 was confirmed as α -orientated due to NOESY correlations with 3H-19 and H-9. NOESY correlations between H-13 and overlapping signals at δH 1.97 (m), 3H-18 and H-14 suggested a β orientation for H-13 and H-14, which is also consistent with the literature (all 6/8/6 taxanes characterized thus far contain H-13 protons oriented in β position (Wang et al., 2011)). Taken together, the evidence presented here indicates that **9** is a novel metabolite, for which the common name 2-deacetyl baccatin IV is suggested (systematic name: $1\beta,2\alpha$ -dihydroxy- $4\alpha,7\beta,9\alpha,10\beta,13\alpha$ -pentaacetoxy- $5\beta,20$ -epoxytax- 11 -ene).

2.4 Searching Spektraris with NMR data from compound mixtures

The isolation of highly enriched fractions of specific taxanes for NMR spectroscopy requires a significant time commitment. It would thus be desirable to also be able to search with NMR data obtained for mixtures of compounds. To test the utility of the Spektraris-NMR tool for enabling such searches, the ^{13}C -NMR chemical shifts (with a 0.5 ppm tolerance setting) from spectra for six taxanes were combined, three of which differed only in the side-chain at C-3' (taxol (**1**), taxol B (**7**), and taxol C (**8**)), while the remainder had different acetylation patterns on the diterpene core (baccatin III (**2**), baccatin IV (**5**), and 4-deacetyl baccatin IV (**10**)). When the combined dataset was used for Spektraris searching

(“Mixture” was selected for Alignment Scoring), the individual compound records were returned as hits #1 (2), #5 (8), #6 (7), #8 (1), #11 (5), and #26 (10). A second Spektraris search with a combination of ^{13}C -NMR spectroscopic data and chemical formulae (as inferred from HPLC-QTOF-MS data described under 2.3) for the compounds in the mixture produced a result table in which compound records were returned as hits #1 (2), #2 (8), #3 (7), #4 (1), #5 (5), and #7 (10) (Supplementary Table 2), thus demonstrating the remarkable selectivity achievable when integrating search functions uniquely supported in Spektraris-NMR.

3. Concluding remarks

Spektraris-AMT, an online resource for PNP identification based on high resolution HPLC/MS data, was recently described (Cuthbertson et al., 2013). In conjunction with MS/MS spectral searches that are enabled, for example, in MassBank (Horai et al., 2010), the AMT tag approach can be used to interrogate data sets acquired with plant extracts for the presence of known PNPs. Here, Spektraris-NMR is now introduced, which expands these capabilities by allowing users to search NMR spectra for a match with data reported in the literature. The utility is demonstrated by screening metabolite fractions isolated from *Taxus × media* cv. Hicksii cell cultures for known taxane metabolites. The only spectra that did not generate high confidence hits in searches against Spektraris-NMR were obtained with a metabolite later demonstrated to have a novel structure, thereby validating the concept of using spectral databases for de-replication purposes.

The greatest challenge for development of Spektraris-NMR was extraction of relevant spectral data from the literature. Unfortunately, although reporting standards for publications with NMR data have been proposed (Rubtsov et al., 2007), there are no generally accepted guidelines or widely accessible online data repositories. Currently available NMR databases are dependent on submissions from the community and the hard work of a few individuals that interpret and integrate data from the literature. The availability of spectroscopic archives/repositories and encouragement to submit spectra in a common format (even those of known compounds) with journal publications would certainly ameliorate the development of searchable databases significantly.

Tremendous progress has been made in recent years with algorithms to predict NMR spectra based on known structures. However, predictions for highly functionalized PNPs that belong to a larger class of structurally related metabolites are not yet viable. To support further developments in this area, the entire spectral database for taxane diterpenoids (466 spectra), including full spectral interpretations, was submitted to NMRShiftDB (Steinbeck et al., 2003; Steinbeck and Kuhn, 2004). It is anticipated that the data exchange between resources that use complementary approaches for spectral searching will propel the field and result in advancements to further increase the utility for users.

4. Experimental

4.1. Chemicals

9-Anthracene carboxylic acid was purchased from Sigma-Aldrich (St. Louis, MO, USA). CDCl_3 was from Cambridge Isotope Laboratories Inc. (Andover, MA, USA). A mixture of 13 taxane standards was purchased from Hauser Chemical Inc., (Boulder, CO, USA). Solvents for metabolite extraction were of analytical reagent grade. HPLC-grade solvents were employed for metabolite purification. The HPLC-QTOF-MS analysis of taxanes was performed with HPLC/MS-grade reagents.

4.2. HPLC-QTOF-MS Analysis

HPLC-QTOF-MS was performed with a 1200 series HPLC system (G1379B degasser, G1312B binary pump, G1367C autosampler, G1316B column heater, G1315C diode array detector, and G1310A isocratic pump) interfaced with a 6520 series QTOF mass spectrometer (Agilent Technologies, Santa Clara, CA, USA). Separations were achieved using a Zorbax SBAq Rapid resolution HT column ($2.1 \times 100 \text{ mm} \times 1.8 \mu\text{M}$) with a Zorbax SB-C8 guard column ($2.1 \times 30 \text{ mm} \times 3.5 \mu\text{M}$) (Agilent Technologies, Santa Clara, CA, USA). The mobile phase was composed of H_2O containing 0.1 % (v/v) HCO_2H (solvent A) and CH_3CN containing 0.1 % (v/v) HCO_2H (solvent B). The solvent composition was 25 % solvent B at injection, which increased linearly to 65 % B over 20 min. At 20.1 min solvent B was increased to 97 % and this composition held for 4 min. The solvent composition was set to return to 25 % solvent B at 24.1 min and the column re-equilibrated for 6 min prior to the next injection. An injection volume of 2.5 μL was used and the flow rate was set to 0.4 mL/min. Chromatograms were recorded at 230 and 254 nm and UV spectra were obtained from 200–400 nm. A reference mass solution containing 250 nM hexakis-(1H, 1H, 3H-tetrafluoropropoxy)-phosphazine (calculated m/z: 922.0098) and 300 nM purine (calculated m/z: 121.0509) in $\text{CH}_3\text{CN}:\text{H}_2\text{O}$ (95:5, v/v) was infused into the eluent at a flow rate of 0.2 mL/min using a T junction fitting located between the diode array detector and the mass spectrometer ion source (with no impact on chromatography). Samples were analyzed by QTOF-MS using an electrospray ionization (ESI) probe operating in negative polarity mode or an atmospheric pressure chemical ionization (APCI) probe in positive mode. The m/z scan range was 50–1,200 amu, the VCAP voltage was set to 3,500 and the fragmentor voltage for MS/MS was 175. The ESI probe was operated with a gas temperature of 325°C, a gas flow of 10 L/min, and a nebulizer pressure of 2.4 bar. The APCI probe was operated with a gas temperature of 325°C, a vaporizer temperature of 350°C, a gas flow of 10 L/min, a nebulizer pressure of 2.4 bar, and a corona value of 15 with positive charge. Solutions containing authentic taxane standards were spiked with anthracene 9-carboxylic acid as an internal standard.

4.3. Extraction and purification of taxanes

Suspension-cultured *Taxus $\times$ media* cv. Hicksii cells (maintained as described in Ketchum et al., 2003) were separated from the culture medium by filtration through a nylon sieve (125 μm mesh size) and directly lyophilized. Cells were weighed, homogenized using a commercial blade-type coffee grinder, and extracted with CH_2Cl_2 (10 mL/g of cells) for 3 h at 23°C. The solvent was evaporated to dryness in vacuo and the residue resuspended in

CH₃CN-H₂O (1:1, v, 132 mg/mL) by vigorous shaking for 30 s (Vortex mixer at highest speed) and sonication in a water bath for 1 h at 23°C. No cell debris was visible. The mixture was then centrifuged glass tubes (4mL) at 500 × g for 5 min (clinical table top centrifuge), which resulted in the formation of two phases. The lower, darker colored, phase contained taxanes and was carefully removed with a Pasteur pipette. Purified taxanes were diluted to a concentration of 0.01-0.1 mg/mL in CH₃CN, and the solution was spiked with 0.01 mg/mL 9-ACA as an internal standard. Each sample was filtered through a 0.22 μM polypropylene membrane prior to analysis. Aliquots of this extract (50 μL per injection) were loaded onto a Discovery HSF5 column (250 × 10 mm × 5 μm; Supelco, Bellefonte, PA, USA) operated at a flow rate of 1 mL/min using CH₃CN:H₂O (42:58; v/v) as isocratic solvent mixture. The HPLC system consisted of an 1100 series G1322A degasser, G1312A binary pump, G1329A autosampler, G1315A diode array detector (monitoring at 230 nm) and G1364A fraction collector (Agilent Technologies, Santa Clara, CA, USA). Following a 3 min delay, 120 fractions were collected over 24 min (12 s per fraction), and the process was repeated until the entire taxane extract was exhausted. Each fraction was evaporated to dryness and the residue stored at -20°C until further processing. An aliquot of each fraction was analyzed by rapid resolution HPLC-QTOF-MS as described below. Fractions with very similar chemical profiles were combined and further purified with a Gemini C18 column (250 × 4.6 mm × 5 μm; Phenomenex, Torrance, CA, USA), operating at a flow rate of 1 mL/min and mobile phases composed of CH₃CN:H₂O (various gradients (depending on properties of metabolites in fraction of interest) from 20:80 to 95:5; v/v) on the HPLC system mentioned above. Fractions were collected every 30 s. An aliquot of each fraction was analyzed by HPLC-QTOF-MS as described above and fractions with identical chemical profiles were combined. CH₃CN was removed *in vacuo* and taxanes were recovered by extraction with EtOAc. The solvent was removed under a stream of N₂ gas and solid taxane residues were stored at -20°C in glass vials. This isolation procedure yielded **1** (0.5 mg), **2** (3.6 mg), **3** (1 mg), **4** (10.6 mg), **5** (6.8 mg), **6** (6.7 mg), **7** (7.4 mg), **8** (15 mg), and **9** (17.6 mg).

4.4. NMR spectroscopy and further characterization of metabolite 9

Routine NMR spectra for the Spektraris-NMR database were acquired on a DRX-400 spectrometer (Bruker BioSpin, The Woodlands, TX, USA) operating at 399.76 MHz for ¹H and 100.52 MHz for ¹³C. Metabolites were dissolved in CDCl₃ and residual solvent was used as an internal reference corrected to 7.26 ppm and 77.16 ppm for ¹H and ¹³C NMR spectroscopy, respectively. The NMR spectrometer temperature was 25°C. ¹H NMR spectra were acquired with a relaxation delay of 5 s, a pulse angle of 45°, and an acquisition time of 3.408 s. ¹³C NMR spectra were acquired with a 1 s relaxation delay, a pulse angle of 45°, and an acquisition time of 1.398 s. 2-Deacetylbaicatin IV (**9**): amorphous colorless solid; MS: APCI(pos) [M+H]⁺ = 611.2687, ESI(neg) [M-H]⁻ = 609.2528; Molecular Formula: C₃₀H₄₂O₁₃; Calculated: xxxxx IR: ν_{max} (KBr) 3471 (m), 1745 (s), 1236 (s) cm⁻¹; For ¹H and ¹³C NMR spectroscopic data, see Table 1 and Supplementary Fig. 1. The full NMR data sets are also available for download at <http://langelabtools.wsu.edu/nmr/>.

4.5. Spektraris-NMR database

A literature search was conducted to find NMR spectra for taxanes bearing a 6/8/6 skeleton. NMR spectral data (together with meta information) were entered into a Microsoft Excel spreadsheets. The data were transcribed from the spreadsheets into an SQL database by a Python script. A web interface for searching unknown spectra against the database was written using the Python Django web framework (<https://www.djangoproject.com/>). The source code for the database and web applications, and spectral data files generated to populate the database are available for download at <http://langelabtools.wsu.edu/nmr/>. Spektraris searches can be based on any combination of: monoisotopic mass, molecular formula, ^1H NMR, and ^{13}C NMR. The first step of the search algorithm is to eliminate from consideration any compound in the database with an monoisotopic mass outside of the tolerance range of the query mass, or a molecular formula not matching the query elemental composition. The query NMR spectral data are then compared against all spectra in the database. The spectroscopic comparisons are modeled as an assignment problem and executed using the Scikit-learn implementation (Pedregosa et al., 2011) of the Hungarian algorithm (Kuhn, 1955). Query peaks (entered as chemical shifts in ppm) are compared to database entries such that the sum of the squared distances between matched peaks is minimized. Any distances further than the tolerance are considered to be equal to the tolerance. After the assignment problem is solved, each query shift assigned to a database shift at a distance closer than the tolerance is considered to be a “matched shift”. Similarity scores (in percent) are calculated by dividing the total number of matched shifts by either the number of query shifts or the number of database shifts, whichever is larger, and multiplying by 100. To account for peak shape or coupling constants in proton spectral comparisons, the distance between two shifts where peak shape or coupling constants are not the same is considered to be equal to the shift distance tolerance (if the data are missing in either the query or the database, this condition does not apply).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by the National Institutes of Health (award numbers 1RC2GM092561-01 (to B.M. Lange) and T32GM083864 (training grant funds supporting S.R. Johnson)), and McIntire-Stennis formula funds from the Agricultural Research Center at Washington State University. The authors would like to thank Ms. Lea Wherland and Ms. Amber Parrish for maintaining *Taxus* cell cultures, Dr. Greg Helms (Center for NMR Spectroscopy, Washington State University) for technical support during the acquisition of NMR spectra, and Stefan Kuhn (University of Leicester, UK) for help with integrating the data into NMRShiftDB.

References

- Appendino G, Cravotto G, Enriù R, Gariboldi P, Barboni L, Torregiani E, Gabetta B, Zini G, Bombardelli E. Taxoids from the roots of *Taxus × media* cv. Hicksii. J Nat Prod. 1994; 57:607–613.
- Banfi D, Patiny L. www.nmrdb.org: resurrecting and processing NMR spectra on-line. CHIMIA. 2008; 62:280–281.
- Barboni L, Gariboldi P, Torregiani E, Appendino G, Gabetta B, Bombardelli E. Taxol analogues from the roots of *Taxus × media*. Phytochemistry. 1994; 36:987–990.

- Chmurny GN, Hilton BD, Brobst S, Look SA, Witherup KM, Beutler JA. ¹H- and ¹³C-nmr Assignments for taxol, 7-epi-taxol, and cephalomannine. *J Nat Prod.* 1992; 55:414–423. [PubMed: 1355110]
- Clerc T, Erni F. Identification of organic compounds by computer-aided interpretation of spectra. *Fortschr Chem Forsch.* 1973; 39:91–107.
- Cragg GM, Newman DJ. Natural products: a continuing source of novel drug leads. *Biochim Biophys Acta.* 2013; 1830:3670–3695. [PubMed: 23428572]
- Cui Q, Lewis IA, Hegeman AD, Anderson ME, Li J, Schulte CF, Westler WM, Eghbalian HR, Sussman MR, Markley JL. Metabolite identification via the Madison Metabolomics Consortium Database. *Nat Biotech.* 2008; 26:162–164.
- Cuthbertson DJ, Johnson SR, Piljac-Zegarac J, Kappel J, Schäfer S, Wüst M, Ketchum REB, Croteau RB, Marques JV, Davin LB, Lewis NG, Rolf M, Kutchan TM, Soejarto DD, Lange BM. Accurate mass-time tag library for LC/MS-based metabolite profiling of medicinal plants. *Phytochemistry.* 2013; 91:187–197. [PubMed: 23597491]
- Davies AN, Lampen P. JCAMP-DX for NMR. *Appl Spectrosc.* 1993; 47:1093–1099.
- Duquesnoy E, Paoli M, Castola V, Bighelli A, Casanova J. Identification of taxanes in extracts from leaves of *Taxus baccata* L using ¹³C-NMR spectroscopy. *Phytochem Anal.* 2009; 20:246–252. [PubMed: 19333914]
- Eldridge GR, Vervoort HC, Lee CM, Cremin PA, Williams CT, Hart SM, Goering MG, O'Neil-Johnson M, Zeng L. High-throughput method for the production and analysis of large natural product libraries for drug discovery. *Anal Chem.* 2002; 74:3963–3971. [PubMed: 12199561]
- Fortineau AD. Chemistry perfumes your life. *J Chem Educ.* 2004; 81:45–50.
- Fukushima A, Kusano M. Recent progress in the development of metabolome databases for plant systems biology. *Front Plant Sci.* 2013; 4:73. [PubMed: 23577015]
- Gunawardana GP, Premachandran U, Burres NS, Whittern DN, Henry R, Spanton S, McAlpine JB. Isolation of 9-dihydro-13-acetylbaccatin III from *Taxus canadensis*. *J Nat Prod.* 1992; 55:1686–1689.
- 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, Tanaka K, 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. [PubMed: 20623627]
- Jones DP, Park Y, Ziegler TR. Nutritional metabolomics: progress in addressing complexity in diet and health. *Annu Rev Nutr.* 2012; 32:183–202. [PubMed: 22540256]
- Ketchum REB, Rithner CD, Qui D, Kim YS, Williams RM, Croteau RB. *Taxus* metabolomics: methyl jasmonate preferentially induces production of taxoids oxygenated at C-13 in *Taxus × media* cell cultures. *Phytochemistry.* 2003; 62:901–909. [PubMed: 12590117]
- Kuhn HW. The Hungarian method for the assignment problem. *Naval Res Logistics.* 1955; 2:83–97.
- Li JW, Vederas JC. Drug discovery and natural products: end of an era or an endless frontier? *Science.* 2009; 325:161–165. [PubMed: 19589993]
- Ludwig C, Easton JM, Lodi A, Tiziani S, Manzoor SE, Southam AD, Byrne JJ, Bishop LM, He S, Arvanitis TN, Günther U, Viant MR. Birmingham Metabolite Library: a publicly accessible database of 1-D ¹H and 2-D ¹H J-resolved NMR spectra of authentic metabolite standards (BML-NMR). *Metabolomics.* 2012; 8:8–18.
- Masui H, Yoshida M. SPECTRA: a spectral information management system featuring a novel combined search function. *J Chem Inf Comput Sci.* 1996; 36:294–298.
- Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay E. Scikit-learn: machine learning in Python. *J Mach Learn Res.* 2011; 12:2825–2830.
- Rojas AC, De Marciano D, Méndez B, de Méndez J. Carbon-13 NMR spectra of taxane-type diterpenes: Oxiranes and oxetanes. *Org Magn Reson.* 1983; 21:257–260.

- Rubtsov DV, Jenkins H, Ludwig C, Easton J, Viant MR, Günther U, Griffin JL, Hardy N. Proposed reporting requirements for the description of NMR-based metabolomics experiments. *Metabolomics*. 2007; 3:223–229.
- Sawada Y, Nakabayashi R, Yamada Y, Suzuki M, Sato M, Sakata A, Akiyama K, Sakurai T, Matsuda F, Aoki T, Hirai MY, Saito K. RIKEN tandem mass spectral database (ReSpect) for phytochemicals: a plant-specific MS/MS-based data resource and database. *Phytochemistry*. 2012; 82:38–45. [PubMed: 22867903]
- Schneider B, Gershenzon J, Graser G, Hölscher D, Schmitt B. One-dimensional ^{13}C NMR and HPLC- ^1H NMR techniques for observing carbon-13 and deuterium labelling in biosynthetic studies. *Phytochem Rev*. 2003; 2:31–43.
- Sénilh V, Blechert S, Colin M, Guénard D, Picot F, Potier P, Varenne P. Mise en évidence de nouveaux analogues du taxol extraits de *Taxus baccata*. *J Nat Prod*. 1984; 47:131–137.
- 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; 27:747–51. [PubMed: 16404815]
- Steinbeck C, Krause S, Kuhn S. NMRShiftDB - constructing a free chemical information system with open-source components. *J Chem Inf Comput Sci*. 2003; 43:1733–1739. [PubMed: 14632418]
- Steinbeck C, Kuhn S. NMRShiftDB - compound identification and structure elucidation support through a free community-built web database. *Phytochemistry*. 2004; 65:2711–2717. [PubMed: 15464159]
- Tulpan D, Léger S, Belliveau L, Culf A, Uperlövi -Culf M. MetaboHunter: an automatic approach for identification of metabolites from ^1H -NMR spectra of complex mixtures. *BMC Bioinformatics*. 2011; 12:400. [PubMed: 21999117]
- Wang YF, Shi QW, Dong M, Kiyota H, Gu YC, Cong B. Natural taxanes: developments since 1828. *Chem Rev*. 2011; 111:7652–7709. [PubMed: 21970550]
- Wishart DS, Tzur D, Knox C, Eisner R, Guo AC, Young N, Cheng D, Jewell K, Arndt D, Sawhney S, Fung C, Nikolai L, Lewis M, Coutouly MA, Forsythe I, Tang P, Shrivastava S, Jeroncic K, Stothard P, Amegbey G, Block D, Hau DD, Wagner J, Miniaci J, Clements M, Gebremedhin M, Guo N, Zhang Y, Duggan GE, MacInnis GD, Weljie AM, Dowlatabadi R, Bamforth F, Clive D, Greiner R, Li L, Marrie T, Sykes BD, Vogel HJ, Querengesser L. HMDB: the human metabolome database. *Nucl Acids Res*. 2007; 35:D521–D526. [PubMed: 17202168]
- Yamamoto O, Someno K, Wasada N, Hiraishi J, Hayamizu K, Tanabe K, Tamura T, Yanagisawa M. An integrated spectral database system including IR, MS, ^1H -NMR, ^{13}C -NMR, ESR and Raman spectra. *Anal Sci*. 1988; 4:233–239.
- Zhang J, Sauriol F, Mamer O, Zamir LO. Taxoids from the needles of the Canadian yew. *Phytochemistry*. 2000; 54:221–230. [PubMed: 10872214]

- Introduction of Spektraris NMR resource for identification of known plant natural products.
- Utility demonstrated by successful identification of taxanes isolated from *Taxus* cell cultures.
- Structure of a taxane, 2-deacetylbaccatin IV, is also determined.
- Spektraris NMR is a valuable tool for dereplication efforts in natural products isolation.

Fig. 1.
Structures of metabolites.

Fig. 2.

Spektraris-NMR database. Flow diagrams of (A) the database development process and (B) the protocol for searching the database. Examples of (C) the data input and (D) data outputformats. For details see text. Acronyms: AMT tag, accurate mass-time tag; LC, liquid chromatography; MFE, molecular feature extraction; MS, mass spectrometry; NMR, nuclear magnetic resonance spectroscopy.

Fig. 3. HPLC-QTOF-MS analysis of metabolite fractions isolated from *Taxus × media* cv. Hicksii cell suspension cultures. (A-I) Total ion currents are shown for fractions containing purified metabolites **1-9** (see Fig. 1 for structures). (J) Mass spectrum of metabolite **9** acquired in APCI(pos) with annotated fragmentation patterns.

Table 1

¹H and ¹³C NMR spectroscopic data for 9 in CDCl₃ (400 MHz)

Position	$\delta(^1\text{H})$, mult	J (Hz)	$\delta(^{13}\text{C})$, mult [*]	HMBC (C→H)	NOESY
1 (OH)	2.53 (s)		42.60	16, 17, 1-OH, 13	
2	4.11 (dd)	5.5, 5.4	72.95	14 α , 1, 5	19, 9
2 (OH)	2.84 (o. br d)	6.0			
3	2.84 (o. d)	5.2	47.23		6 α
4			82.61	20b, 5	
5	4.96 (d)	8.3	83.74	6 α , β , 20a	6 α
6 α	2.48 (m)		34.77	7	7, 5
6 β	1.89 (m)				
7	5.49 (dd)	9.8, 7.8	72.24	19, 6 α , β , 5, 9	3, 6 α
8			45.87	19, 6 α , 2, 9	
9	5.84 (d)	11.3	75.32	19, 10	19
10	6.17 (d)	11.3	70.54	9	16, 7
11			133.69	16, 17, 9	
12			141.10	14 α , 13	
13	6.16 (m)		69.89	14 α	14 β , 18
14 α	2.14 (m)		35.58	2, 13, 1-OH	
14 β	1.97 (o. m)				13
15			78.50	17, 16, 14 α , 2, 1-OH	
16	1.25 (s)		28.26	17	10
17	1.65 (s)		22.48	16	
18	1.97 (o. s)		21.11 [†]		13
19	1.60 (s)		12.94	9, 7	9, 2
20a	4.63 (dd)	8.9, 0.9	77.85		
20b	4.58 (d)	8.9			
OAc	2.15 (s)		170.67 [‡]		
			22.92		
OAc	2.15 (s)		169.05 [‡]		
			21.36		

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Position	$\delta(^1\text{H})$, mult	J (Hz)	$\delta(^{13}\text{C})$, mult [*]	HMBC (C→H)	NOESY
OAc	2.10 (s)		170.14		
			21.60		
OAc	2.08 (s)		170.63		
			20.96		
OAc	1.97 (o. s)		169.11		
			15.11 [†]		

o. = overlap.

^{*} Multiplicity assigned by DEPT spectral analysis.[†] Signals interchangeable.[‡] Signals interchangeable.