Dereplication of Natural Products using Minimal NMR Data Inputs

Journal:	*Organic & Biomolecular Chemistry*
Manuscript ID	OB-REV-08-2015-001713
Article Type:	Review Article
Date Submitted by the Author:	14-Aug-2015
Complete List of Authors:	Moser, Arvin; ACD/Labs, Williams, Russell; Sequoia Sciences Inc., O'Neil-Johnson, Mark; Sequoia Sciences Inc., Williams, Antony; ChemConnector Inc., Wheeler, Patrick; ACD/Labs, Pol, Rostisalv; ACD/Labs,

Organic & Biomolecular Chemistry

ARTICLE

Dereplication of Natural Products using Minimal NMR Data Inputs

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

www.rsc.org/

Russell B. Williams,[a] Mark O'Neil-Johnson,[a] Antony J. Williams,[b] Patrick. Wheeler,[c] Rostislav Pol[c] and Arvin Moser[c*]

A strategy for the dereplication of a complete or a partial structure using ^{1}H NMR, ^{1}H-^{13}C HSQC and ^{1}H-^{1}H COSY spectral data, a molecular formula composition range and structural fragments against a massive database of about 22 million compounds is considered. As the increasing availability of public online databases containing natural products continues to grow the potential of utilizing these resources for dereplication purposes increases. This work examines approaches for NMR dereplication of natural products and includes a comparison with approaches for molecular formula and mass-based dereplication. The strategy is an application of computer-assisted structure elucidation using ACD/Structure Elucidator and data obtained from the ChemSpider database hosted by the Royal Society of Chemistry.

Introduction

Dereplication is the process of testing samples of mixtures that are active in a screening process, so as to recognize and eliminate substances that have already been characterized.[1] The process is directed by a minimal set of analytical data inputs used to search across a database of known materials. Such inputs generally include molecular formula (generally obtained by accurate mass measurement), λ_{max} from UV-Vis spectroscopy, chemical shifts from NMR, and molecular fragments evident from NMR spectroscopy and mass spectrometry.[2-3] In this work we investigate a strategy for how to utilize a large database of almost 22 million diverse chemical compounds for the dereplication of natural products (NPs).

Such methods are becoming increasingly important because as the number of reported NPs increases, it is vital to have an efficient method for directing discovery efforts. This is especially true in drug discovery efforts where the expense of *de novo* isolation and structure elucidation of known compounds is prohibitive. The most common method of dereplication is the use of mass spectrometry, MS.[4] However, MS-based methods can lead to the misidentification of compounds due to differences in ionization and multiple molecular entities having the same molecular formula (MF).

The approach we outline focuses on ^{1}H and ^{13}C chemical shifts from ^{1}H NMR, ^{1}H-^{1}H COSY and ^{1}H-^{13}C HSQC NMR spectra, molecular formula composition ranges and structural fragment information and relates this to monoisotopic mass and molecular formula. Additionally, we examine strategies for identifying not only compounds that are present, but also those that are not present in the database.

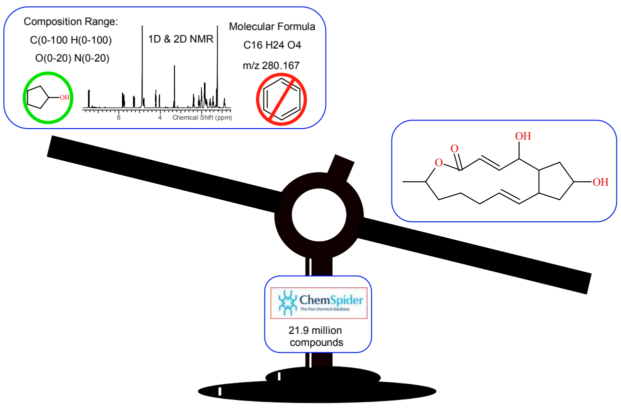


Figure 1. The dereplication model utilizes the ChemSpider database filtered by data such as ^{1}H and ^{13}C chemical shifts, molecular formula, composition range, mass and user fragments.

Experimental

Materials and Methods

The Royal Society of Chemistry is the host of the ChemSpider database,[5] a public resource hosting over 35 million chemical compounds of both synthetic and natural origin - a small proportion of these being NPs and estimated to be about 0.2% of the collection. A subset of the ChemSpider

[a] Sequoia Sciences, Inc., 1912 Innerbelt Business Center Drive, St. Louis, MO 63114, USA
[b] ChemConnector Inc., 904 Tamaras Circle, Wake Forest, NC, 27587, USA
[c] Advanced Chemistry Development, Toronto Department, 8 King Street E, 107, Toronto, Ontario, M5C 1B5, Canada
*Email: Arvin Moser (arvin@acdlabs.com)

Electronic Supplementary Information (ESI) available: Experimental ^{1}H NMR, ^{1}H-^{13}C HSQC (or HMQC), ^{1}H-^{1}H COSY are provided for the 18 compounds. Full search results are provided. See DOI: 10.1039/x0xx00000x database, made up of small organic molecules was prepared: no multiple component compounds, formulae limited only to those containing C, H, O, N, S, P, F, Cl, Br and I from the public ChemSpider database. These files, containing a total of ca. 22 million records, were merged with the data contained within the ACD/Structure Elucidator software.[6] The database was setup to be searched for both synthetic and natural compounds to develop a strategy regarding the use of a massive database. ^{1}H, ^{13}C, ^{15}N, ^{19}F, and ^{31}P NMR chemical shifts were predicted utilizing ACD/C&H NMR Predictors and ACD/XNMR Predictors.[6] The structures were also used to generate the monoisotopic masses and the molecular formulae. Each structure contained their respective ChemSpider ID# to link back to the ChemSpider source page on the website. External sources offered additional information on the respective compound.[7-8] For all work discussed here ACD/Structure Elucidator v.14.02 Nov 21, 2014 running on a Windows 7 64 bit, RAM 8 GB, Dual Core 2.67 GHz computer platform was used. The local version of the ChemSpider database is included at no charge with ACD/Structure Elucidator.

Compounds were isolated using flash chromatography, preparative HPLC and semi-preparative HPLC as previously described.[9] The samples (15-50 μg) were dissolved in 13 μL of solvent. All NMR data reported in this work were obtained on a Bruker Avance 600 MHz spectrometer equipped with a Bruker BioSpin TCI 1.7 mm MicroCryoProbe. All NMR spectra were acquired in CD_3OD or DMSO-d_6 solvent. The ^{1}H NMR spectra had a minimum S/N of 30 with a purity of 80% or higher. These compounds were elucidated by hand from previous work using ^{1}H NMR, ^{1}H-^{13}C HSQC (or HMQC), ^{1}H-^{1}H COSY, ^{1}H-^{1}H ROESY ,^{1}H-^{13}C HMBC and MS data. The identification of compound #6 relied solely on ^{1}H NMR, ^{1}H-^{1}H COSY, MS data and information on structural analogues.

All NMR spectra were peak-picked by hand. Impurities such as solvent signals, impurities and artefacts were not included in the search list. Peaks obscured by solvents, impurities or artefacts were only used if they were clearly discernible. Overlapping multiplets were picked as one peak. The ^{1}H chemical shifts were used in the search but the coupling pattern and coupling constants were not used.

The NMR searches relied solely on the ^{1}H or ^{13}C chemical shifts with a default looseness factor to influence the number of matches. The looseness allowed the number of matching signals to be defined so that not all experimental query signals needed to match. In addition, the chemical shifts were allowed some looseness criterion to address potential deviations between the predicted shifts in the database and the experimental shifts from the spectra. The ^{13}C search used the ^{13}C multiplicity (i.e. CH, CH_2 and CH_3) as determined using the information from the experimental ^{1}H integrals and the correlations from the ^{1}H-^{13}C HSQC data.

The key filter settings for spectral data searching were:

1. Reject structures with a ^{1}H and ^{13}C NMR deviation of more than 0.15 and 2.0 ppm, respectively,
2. Allow for a lack of signals in the ‘full’ structure of up to 6 signals for ^{1}H and 2 signals for ^{13}C chemical shifts (*i.e.* all experimental chemical shifts from the NMR spectrum do not need to match up with all the shifts from the hit structure),
3. Allow an excess of 0 signals for ^{1}H and 10 signals for ^{13}C chemical shifts in the hit structure (*i.e.* when set to 0, all the shifts from the hit structure do need to match all the experimental chemical shifts from the NMR spectrum), and
4. Ignore the peak intensities during the search.

For point #3, in three cases of the ^{1}H benchmark test set, the search results turned up no hits and so this value was increased to 2, 6 and 8 for structures #7, 8 and 16, respectively. For the ^{13}C benchmark test set, this value was increased to 15, 15 and 20 for structures #7, 8 and 16, respectively The three cases had multiple diastereotopic CH_2 protons or had an asymmetrical C_2 axis.

The ^{13}C chemical shifts were obtained from the respective ^{1}H-^{13}C HSQC (or HMQC) experiment. In addition, the lower boundary of the composition range was based on the carbon count from the HSQC spectrum. The maximum carbon count set was limited by a reasonable guess of the upper limit.

The benchmark test set used 18 known compounds, 16 of which were found to be present in the ChemSpider database. For the 2 compounds not present in ChemSpider, a substructure good list and bad list was employed to reduce the number of hits. The good list was chosen due to distinct and obvious signals in the NMR spectra. Likewise, the bad list eliminated hits that were not consistent with the NMR spectra. The substructures were pieced together using ^{1}H chemical shifts, coupling patterns and integrals and COSY correlations.

Results and Discussion

As the number of identified NPs increases we believe that it may be of value to develop strategies to facilitate the dereplication process through available databases. While there are commercial databases of value (for example, Marinlit (27,589 records)[10] for marine NPs, Antibase (42,950 records)[11] for microorganisms and higher fungi materials, and the Dictionary of Natural Products (DNP) (>265,000 records)[12]) for a broad collection, none of these provide structural collections in a format that can be integrated into the ACD/Structure Elucidator software. RSC’s ChemSpider database is free to access and contains tens of millions of compounds from various sources and, as a result of other collaborations between RSC and ACD/Labs regarding the PharmaSea project[13] access was provided to an appropriate subset of data from ChemSpider. The identity of an unknown compound can be divided into one of three categories following a database search: known (spectral and/or structural data exists in a database), partially known (an analogue(s) exists in a database), or novel (a unique structure that shares little similarity to any compound in a database).

This journal is © The Royal Society of Chemistry 20xx

For a number of known NPs, data associated with 16 compounds were examined and used to search across the ChemSpider database (see Figure 2). The collection of compounds ranged in molecular weight from 200 to 400 Da.

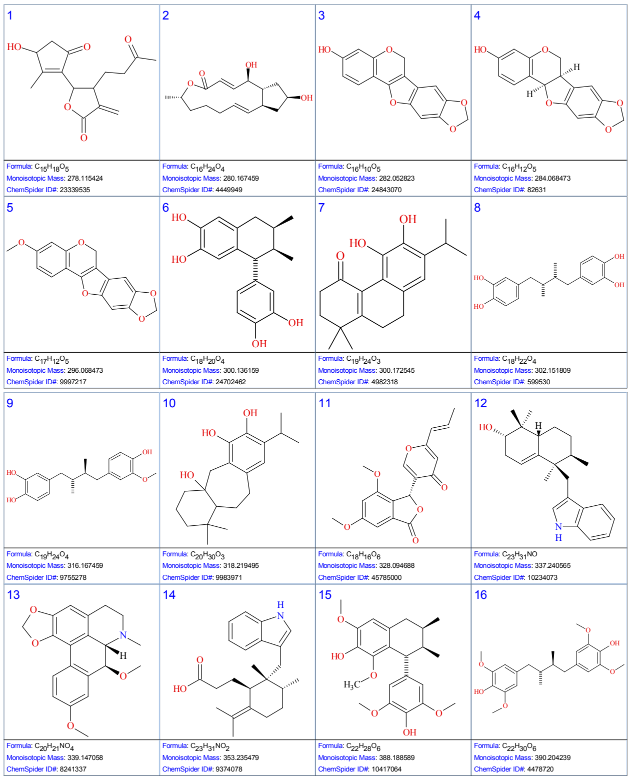

1
Formula: $C_{15}H_{18}O_5$
Monoisotopic Mass: 278.115424
ChemSpider ID#: 23339535
2
Formula: $C_{16}H_{24}O_4$
Monoisotopic Mass: 280.167459
ChemSpider ID#: 4449949
3
Formula: $C_{16}H_{10}O_5$
Monoisotopic Mass: 282.052823
ChemSpider ID#: 24843070
4
Formula: $C_{16}H_{12}O_5$
Monoisotopic Mass: 284.068473
ChemSpider ID#: 82631
5
Formula: $C_{17}H_{12}O_5$
Monoisotopic Mass: 296.068473
ChemSpider ID#: 9997217
6
Formula: $C_{18}H_{20}O_4$
Monoisotopic Mass: 300.136159
ChemSpider ID#: 24702462
7
Formula: $C_{19}H_{24}O_3$
Monoisotopic Mass: 300.172545
ChemSpider ID#: 4982318
8
Formula: $C_{18}H_{22}O_4$
Monoisotopic Mass: 302.151809
ChemSpider ID#: 599530
9
Formula: $C_{19}H_{24}O_4$
Monoisotopic Mass: 316.167459
ChemSpider ID#: 9755278
10
Formula: $C_{20}H_{30}O_3$
Monoisotopic Mass: 318.219495
ChemSpider ID#: 9983971
11
Formula: $C_{18}H_{16}O_6$
Monoisotopic Mass: 328.094688
ChemSpider ID#: 45785000
12
Formula: $C_{23}H_{31}NO$
Monoisotopic Mass: 337.240565
ChemSpider ID#: 10234073
13
Formula: $C_{20}H_{21}NO_4$
Monoisotopic Mass: 339.147058
ChemSpider ID#: 8241337
14
Formula: $C_{23}H_{31}NO_2$
Monoisotopic Mass: 353.235479
ChemSpider ID#: 9374078
15
Formula: $C_{22}H_{28}O_6$
Monoisotopic Mass: 388.188589
ChemSpider ID#: 10417064
16
Formula: $C_{22}H_{30}O_6$
Monoisotopic Mass: 390.204239
ChemSpider ID#: 4478720

Figure 2. The list of 16 known compounds used to analyse the performance of the dereplication process. The molecular formulae, monoisotopic masses and ChemSpider ID number are included. Structures 2 and 11 are from fungi. All of the others are plant NPs.

Various input data can be used to filter down a search result including molecular formula (MF), direct mass value extracted from a mass spectrum, composition range (full or limited), substructure(s) and associated NMR data. A series of different approaches were undertaken. When the mass (or molecular formula) were not known it was possible to perform searches across molecular composition ranges. For this work it was assumed that only the elements C, H, O and N were present in the compounds based on the known composition of the benchmark test set. Two separate composition profiles were tested. The full composition (fC) listed the potential composition ranges for a number of elements with the broad ranges default set at C0-100 H0-100 O0-20 N0-10. The limited composition (lC) was set at C10-30 H10-40 O0-15 N0-5 for structures with less than 25 H atoms and C10-30 H25-40 O0-15 N0-5 for structures with more than 25 atoms.

The complete list of approaches were:

1. m/z search with fC and lC and with 0, 1 and 2 substructures
2. MF with 0, 1 and 2 substructures
3. NMR with MF
4. NMR with m/z and with fC and lC
5. NMR with fC and lC and with 0, 1 and 2 substructures

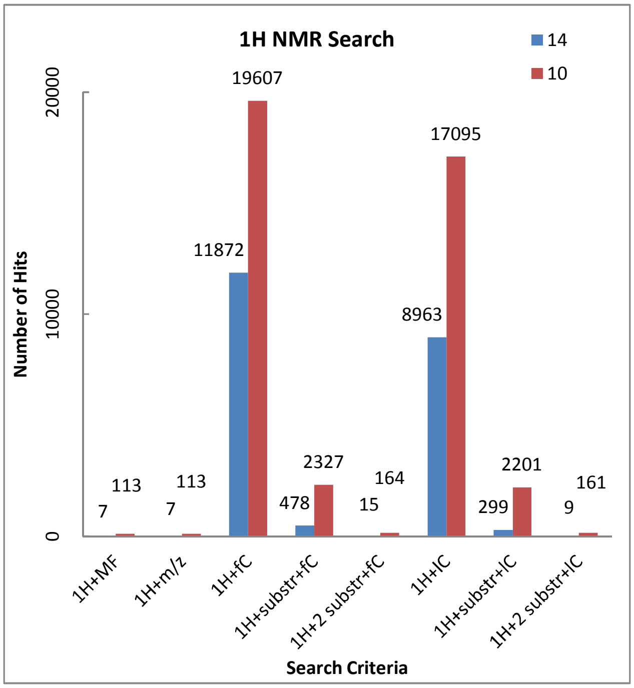


Figure 3. The bar graph of ^1H NMR spectrum search results for NPs 14 and 10 combined with other criteria: MF = molecular formula, m/z tolerance = +/- 0.001 Da, fC = full composition (C0-100 H0-100 O0-20 N0-10), lC= limited composition (C10-30 H25-40 O0-15 N0-5) and substr = substructure.

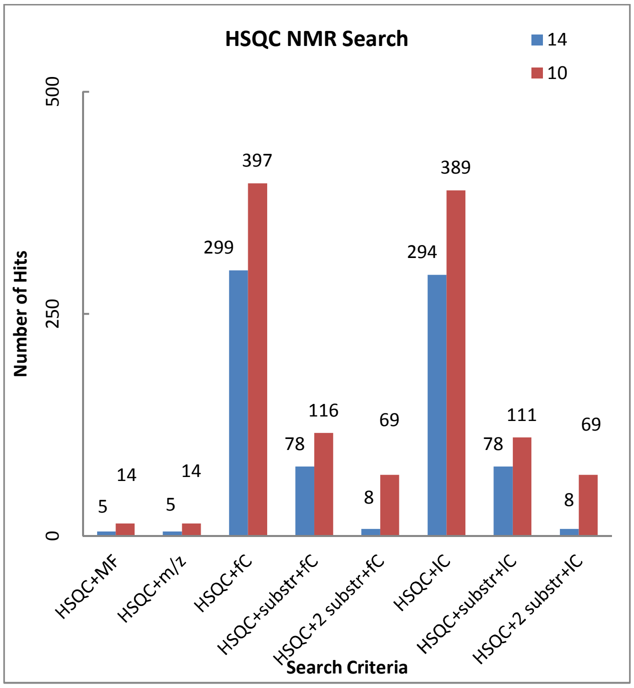


Figure 4. The bar graph of HSQC NMR spectrum search results for NPs 14 and 10 combined with other criteria: MF = molecular

formula, m/z tolerance = +/- 0.001 Da, fC = full composition (C0-100 H0-100 O0-20 N0-10), lC= limited composition (C10-30 H25-40 O0-15 N0-5) and substr = substructure.

As an example of the type of results obtained with these approaches Figures 3 and 4 show the search results for two known NPs using ^1H (from 1D ^1H NMR experiment) and ^{13}C (from HSQC experiment) chemical shifts in combination with various input data. For ^1H NMR search, a reasonable number (<200 hits) of hits are obtained when a ^1H NMR and MF, m/z or two suggested substructures are used to constrain the results. The ^{13}C search results were fairly reasonable without any substructure constraints. For both the ^1H and ^{13}C search results with full composition and limited composition constraints (NMR+fC+substr and NMR+lC+substr), the query compound was ranked in the top 3 for 10 NPs and in the top 20 for 13 NPs. The search times for either ^1H(or HSQC)+MF and ^1H(or HSQC)+m/z were <1 minute and <4 minutes, respectively. The research groups of Bradshaw[3], Lang[14], Bitzer[15] and Johansen[16] used NMR data (with and without MS data) for dereplication with results of less than 10 hits. The few number of hits was due to a query database of about 100 times smaller in number of records than the ChemSpider database.

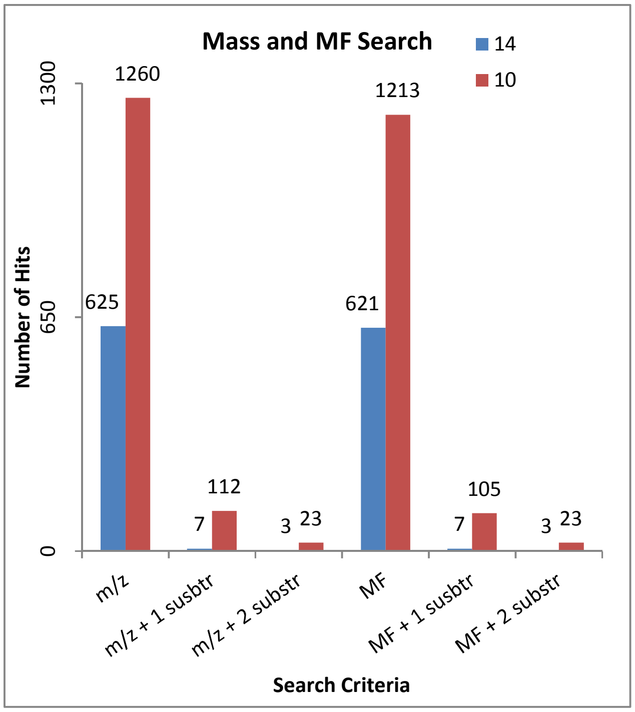
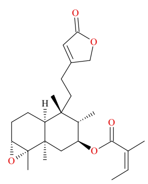


Figure 5. Bar graph of m/z and MF search results for NPs 14 and 10. MF = molecular formula, m/z tolerance = +/- 0.001 Da with a composition (C0-100 H0-100 O0-20 N0-10) and substr = substructure.

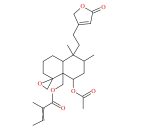
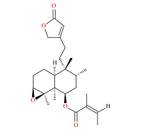
The search results obtained using m/z and MF show a limited and reasonable number of hits when combined with additional data. From Figure 5, the search results of compound #14 based on m/z and MF produced 625 hits and 621 hits, respectively. The high hit count for only m/z and MF are impractical to sort through. The average search times for individually m/z and MF are ~3 minutes and 15 seconds, respectively.

While mass spectrometry can be an incredibly sensitive and fast technique, the molecular mass (or extracted MF) alone is hardly distinctive in terms of dereplication, especially considering the number of chemically appropriate compounds that can exist for a molecular formula. MSn data can be very valuable but it too can be problematic for dereplication work.[17-18] These issues can include ionization and fragmentation problems, selecting the correct m/z and correctly determining if the pseudo-molecular ion is $[M+H]^+$, $[M+Na]^+$, *etc*. are also not necessarily trivial. Identification of an unknown by MS data is further complicated when the mass does not match any compound stored in a database.[19] In those cases, additional structural information is needed and NMR offers a great way to obtain this information and differentiate stereoisomers. Furthermore, with modern instrumentation it is possible to obtain high-quality data in a few minutes using just a few μgs of material.

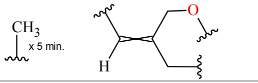
It is important to note that all compounds are initially unknown until a data analysis is attempted. Two new NPs were examined where similar but not identical compounds exist in the database. The ^1H NMR search results of 2 new compounds are illustrated in Figures 6 and 7. The top 2 hits, ranked by (d_{E}(1H)), were based on a ^1H NMR and ^1H-^1H COSY spectra. The Good List were a list of substructures that must be present whereas the Bad List represented substructures that could not be present. The ^1H NMR deviation (d_{E}(1H)) is the overall weighted difference between the experimental chemical shifts and the predicted shifts. That is, a value close to zero is good. The top 2 hits have >78% similarity based on Tanimoto similarity coefficient. The average search time was 13 minutes.



CS ID# Not in ChemSpider
MF C25 H36 O5
Mmi 416.256 Da

Rank Order	Structure Hits	CSID#	Formula	Mmi	d_{E}(1H)
1		433194	$C_{27}H_{38}O_7$	474.261	0.053
2		10397202	$C_{25}H_{36}O_5$	416.256	0.059

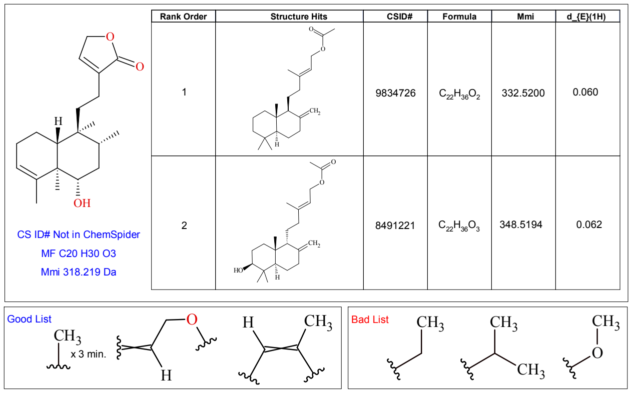
Good List



Bad List

NONE

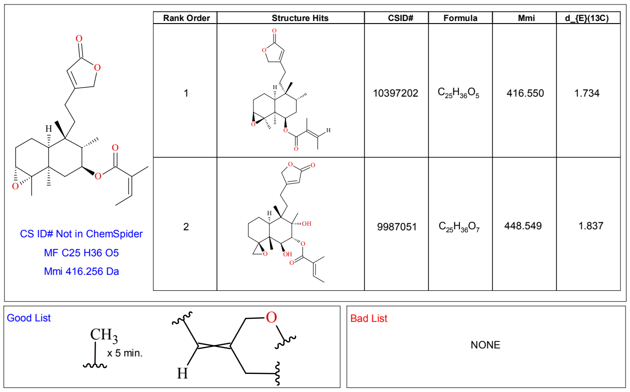
Figure 6. The top 2 hits for the query structure searched by ^1H chemical shifts from the ^1H NMR spectrum. The hits were ranked by the ^1H NMR deviation (d_{E}(1H)) and filtered by a lC = limited composition (C10-30 H25-40 O0-15 N0-5) and the Good List and Bad List. The Good List was determined from the ^1H chemical shifts, ^1H integrals and ^1H-^1H COSY correlations.



Rank Order	Structure Hits	CSID#	Formula	Mmi	d_{E}(1H)
1		9834726	$C_{22}H_{36}O_2$	332.5200	0.060
2		8491221	$C_{22}H_{36}O_3$	348.5194	0.062

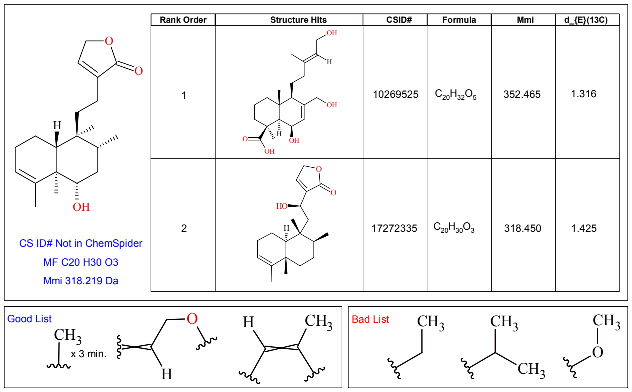
Figure 7. The top 2 hits for the query structure searched by ^{1}H chemical shifts from the ^{1}H NMR spectrum. The hits were ranked by the ^{1}H NMR deviation (d_{E}(1H)) and filtered by a limited composition (C10-30 H25-40 O0-15 N0-5) and the Good List and Bad List. The Good and Bad Lists were determined from the ^{1}H chemical shifts, ^{1}H integrals and ^{1}H-^{1}H COSY correlations.

Figures 8 and 9 illustrate the ^{13}C NMR search results of 2 partially known compounds. The top 2 hits, ranked by (d_{E}(13C)), are based on ^{13}C chemical shifts from a ^{1}H-^{13}C HSQC spectrum. The ^{13}C NMR deviation (d_{E}(13C)) is the overall weighted difference between the experimental chemical shifts and the predicted shifts. Based on the Tanimoto coefficient, the top hits have >80% similarity. The average search time was 13 minutes.



Rank Order	Structure Hits	CSID#	Formula	Mmi	d_{E}(13C)
1		10397202	$C_{25}H_{36}O_5$	416.550	1.734
2		9987051	$C_{25}H_{36}O_7$	448.549	1.837

Figure 8. The top 2 hits for the query structure searched by ^{13}C chemical shifts extracted from the ^{1}H-^{13}C HSQC spectrum. The hits were ranked by the ^{13}C NMR deviation (d_{E}(13C)) and filtered using a limited composition (C10-30 H25-40 O0-15 N0-5) and the Good List and Bad List. The Good List was determined from the ^{13}C chemical shifts from the HSQC correlations, ^{1}H coupling patterns and integrals.



Rank Order	Structure HIts	CSID#	Formula	Mmi	d_{E}(13C)
1		10269525	$C_{20}H_{32}O_5$	352.465	1.316
2		17272335	$C_{20}H_{30}O_3$	318.450	1.425

Figure 9. The top 2 hits for the query structure searched by ^{13}C chemical shifts from the ^{1}H-^{13}C HSQC spectrum. The hits were ranked by the ^{13}C NMR deviation (d_{E}(13C)) and filtered by a limited composition (C10-30 H25-40 O0-15 N0-5) and the Good List and Bad List. The Good and Bad Lists were determined from the ^{13}C chemical shifts from the HSQC correlations, ^{1}H coupling patterns and integrals.

Conclusions

NMR dereplication is a powerful screening process for unknown compounds commonly found in nature but certainly the approach outlined here is also more generally applicable to other types of chemistry, especially considering the generality of the ChemSpider database. Dereplication can be directed by a minimal set of inputs while searching across a chemical database. Such inputs can include an m/z peak from a mass spectrum, ^{1}H or ^{13}C chemical shifts from NMR spectra, a molecular formula, a fragment, a composition range, *etc*. The utilization of a large database of almost 22 million records, while useful, can produce a large number of hits depending on the inputs. Regardless, the NMR dereplication process was effective and successful in identifying both old (found in ChemSpider) and new (not found in ChemSpider) compounds. The ^{1}H search can rely on 1D ^{1}H NMR supported by fragments from a COSY experiment. The ^{13}C search can rely on 1D ^{1}H NMR and HSQC experiments supported by fragments from a COSY experiment. For the new compounds, the top hits have a similarity of >78% based on the Tanimoto coefficient. The average search times for ^{1}H and ^{13}C NMR searching were both 13 minutes.

While it is possible to benefit from using a large database for dereplication, results would definitely be improved if a restricted and focused dataset of NPs only could be used. The ChemSpider database is sourced from many hundreds of data sources including government databases, chemical vendors, publications and patents and the vast majority of the data are therefore not NPs. What would be most ideal is to extract a NPs collection from the database for more direct and applied usage. In a similar way, the approach outlined here could be beneficially used for metabolite identification or pesticides analysis with specific subsets from the ChemSpider database. Furthermore, additional work can be extended to crude samples and mixtures to see the impact of sample purity on the dereplication process.

Acknowledgements

The authors thank Sergey Golotvin for assistance with the software, Joseph C. DiMartino for assistance in preparing the databases and Brent G. Pautler for assistance with the references.

References

1 Webster's Online Dictionary, http://www.webster-dictionary.org/definition/dereplication, Accessed February 2015.
2 C. Steinbeck, *Nat. Prod. Rep.* 2004, **4**, 512.
3 J. Bradshaw, D. Butina, A.J. Dunn, R.H. Green, M. Hajel, M.M. Jones, J.C. Lindon and P.J. Sidebottom, *J. Nat. Prod.* 2001, **64**, 1541.
4 T. Fink, and J.-L. Reymond, *J. Chem. Inf. Model.* 2007, **47**, 342.
5 Royal Society of Chemistry: ChemSpider http://www.chemspider.com/, Accessed May 2015.
6 *ACD/Structure Elucidator, Version 14.0* Advanced Chemistry Development, Inc., Ontario, Canada; 2014.*ACD/CHNMR & XNMR Predictors, Version 14.0* Advanced Chemistry Development, Inc., Ontario, Canada; 2014. http://www.acdlabs.com/products/com_iden/elucidation/struc_eluc/
7 J.L. Little, A.J. Williams, A. Pshenichnov and V. Tkachenko, *J. Am. Soc. Mass Spectrom.* 2012, **23**, 179.
8 C.J. Henrich and J.A. Beutler, *Nat. Prod. Rep.* 2013, **30**, 1284.
9 J.-F.Hu, E. Garo, H.-D. Yoo, P.A. Cremin, L. Zeng, M.G. Goering, M. O'Neil-Johnson and G.R. Eldridge, *Phytochem. Anal.* 2005, **16**, 127-133; G.R. Eldridge, H.C. Vervoort, C.M. Lee, P.A. Cremin, C.T. Williams, S.M. Hart, M.G. Goering, M. O'Neil-Johnson and L. Zeng, *Anal. Chem.* 2002, **74**, 3963.
10 Marinlit April 21, 2015: a database of the marine natural products literature http://pubs.rsc.org/marinlit/
11 Wiley Antibase 2014: The Natural Compound Identifierhttp://ca.wiley.com/WileyCDA/WileyTitle/productCd-3527338411.html
12 Chapman & Hall/CRC Chemical Database: Dictionary of Natural Products (DNP) April 2015 http://dnp.chemnetbase.com/
13 PharmaSea http://www.pharma-sea.eu/, Accessed May 2015.
14 G. Lang, N.A. Mayhudin, M.I. Mitova, L. Sun, S. van der Sar, J.W. Blunt, A.L.J. Cole, G. Ellis, H. Laatsch and M.H.G. Munro, *J. Nat. Prod.*, 2008, **71**, 1595.
15 J. Bitzer, B. Kopcke, M. Stadler, V. Hellwig, Y. Ju, S. Seip, and T. Henkel, *Chimia* 2007, **61**, 332.
16 K.T. Johansen, S.G. Wubshet and N.T. Nyberg, *Analyt. Chem.* 2013, **85**, 3183.
17 K.F. Nielson and T.O. Larsen, *Front. Microbiol.*, 2015, **6**, 71.
18 A. Vaniya and O. Fiehn, *TrAC*, 2015, **69**, 52.
19 J. Hubert, S. Collet, S. Purson, R. Reynaud, D. Harakat, Ag. Martinez, J.-M. Nuzillard and J.-H. Renault, *J. Nat. Prod.* 2015, **78**, 1609.