

# Letter to the Editor: Importance of citing full details of spectral library and search program

**O. David Sparkman**

Adjunct Professor of Chemistry and Director of the Pacific Mass Spectrometry Facility, Chemistry Department, University of the Pacific, Stockton, CA, USA. E-mail: [osparkma@pacific.edu](mailto:osparkma@pacific.edu)

A collection of mass spectra of known compounds and the program used to search a spectrum of an unidentified compound are two different entities. The results obtained depend on both. Two different compounds can be found as the first Hit when a search of the same spectrum is performed using the NIST Mass Spectral (MS) Search Program (National Institute of Standards and Technology, Gaithersburg, MD, USA) to search the NIST/EPA/NIH EI Mass Spectral Library (NIST 20) or the *Wiley Registry of Mass Spectral Data*, 12<sup>th</sup> Edn (John Wiley & Sons, Hoboken, NJ, USA). Two different compounds can be found as the first Hit when the NIST/EPA/NIH EI Mass Spectral Library (NIST 20) is searched using the NIST MS Search Program v.2.4 or the Probability Based Matching (PBM) search program, which is a part of the Agilent Technologies (Santa Clara, CA, USA) ChemStation Data Analysis program. Even when the same mass spectrum is searched against two different versions of the NIST/EPA/NIH EI Mass Spectral Library (NIST 14 and NIST 20) using the same search program (NIST 20's MS Search v.2.4) two different compounds can be found as the first Hit.

There is not one root cause for these differences. In the example of a difference between the search of the NIST 14 and NIST 20 EI libraries using the same search program, it can be as simple as the addition of the spectrum of the

unidentified compound to the library between editions. Spectra for approximately ~60K new compounds were added between these two releases. Or, the different first Hits could be due to complex search algorithm differences in the case of the NIST MS Search program vs the PBM search of ChemStation.

The point is that it is not only necessary to be specific about the publisher of the library but also the edition of the mass spectral library as well as the search program used. Because of my involvement with NIST, I have subscribed to two search strings in Google Scholar Alerts since 5 October 2014 ([NIST Mass Spectral Database] and [nist "mass [spectral | spectra | spectrum]"]). During this period, I have received list of citations for both of these strings every two to five days. Each list contains 5–20 citations to articles that make mention of one of the two NIST Mass Spectral Libraries (NIST/EPA/NIH Library of EI Spectra or the NIST Tandem Library of Production Mass Spectra). Many of these citations also reference the *Wiley Registry* and sometimes other smaller libraries like the flavor and fragrance EI library of Robert P. Adams (Diablo Analytical, Antioch, CA, USA), the Maurer/Pfleger/Weber *Mass Spectral Library of Drugs/Poisons/Pesticides/Pollutants and Their Metabolites*, 2011 Edn (John Wiley & Sons), the Designer Drugs 2020 EI mass spectral library by Peter Rösner (John Wiley & Sons), *SWGDRUG Mass Spectral*

*Library* etc., to mention just some of the many small mass spectral libraries.

Too often citations to the search of a mass spectral library are simply. "The identity of unknowns was confirmed using the NIST library." Unfortunately, not more than 20% of the articles cite the version (edition) of the mass spectral library, and nearly every article is missing any mention of the program used to search the library. When someone purchases an instrument, they will purchase it with a mass spectral library. Today, this is truer of gas chromatography/mass spectrometry (GC/MS) instruments than tandem mass spectrometers used with liquid chromatography, but with the introduction of the NIST Tandem Library and the fact that ThermoFisher is providing a copy of that Library with every tandem instrument that they sell, this will change.

It is also unfortunate that people are publishing an article in 2020 in which the NIST EI library released in 1998 is being used.

As these instrument age, the researchers using them, especially those that are not mass spectrometrists, don't realise that easily and inexpensive upgradable tools, like libraries of mass spectra and search programs that can greatly expand their existing instrument's usefulness, are available.

The main reason journal article authors need to include not only the name and edition of the mass spectral library

used to confirm or identify unidentified compounds but also which programs were used to search these libraries, is so their reader can evaluate the validity of the results.

It is not enough to say, "The identity of the components was confirmed by searching their mass spectra, separately, against the NIST and Wiley libraries". That sentence should read, "The identity of the components was confirmed by searching their mass spectra, separately against the NIST 17 and Wiley 8 libraries using the internal library search algorithm for the Shimadzu GC/MS Solutions V.4.5 software". (The Shimadzu GC/MS Solutions software, like that of most other GC/MS Data Systems, allows only one mass spectral library at a time to be searched.) PerkinElmer, Waters, Agilent's ChemStation (GC/MS and LC/MS) and Agilent's MassHunter (Qual and Quan), Sciex, Bruker and others all have their completely separate proprietary library search routines even through many provide the NIST MS Search Program and libraries in the MS Search format as well as their proprietary format. These proprietary programs and formats can change with changing version of their instrument's software that they accompany; this is why the version number of the data analysis software should be a part of the citation. With the availability of both the NIST MS Search and other

third-party search programs as well as the software's priority search, it is absolutely mandatory that the search software be specified.

It should be remembered that the NIST MS Search Program has both an Identity Search (to be used when it is suspected that a spectrum of the unidentified compound is in the libraries being searched) and a Similarity Search (to be used when it is suspected that there is no spectrum of the unidentified compound in the searched libraries). In the NIST MS Search Program (v.2.4), accompanying NIST 20, there are four different Identity Searches (EI Normal, EI Quick, MS/MS and In-Source HiRes) and five different types of Similarity Searches (EI Hybrid, EI Simple, EI Neutral Loss, MS/MS in EI and MS/MS Hybrid, for use with product-ion mass spectra from atmospheric pressure ionisation produced precursor-ions). Therefore, it is best that not only the version being used is stated but, also, which algorithm is being used when reporting the use of the NIST MS Search Program. Another unique feature of the NIST MS Search, unlike the search routines provided as part of most data analyses' software routines, is that up to 127 different libraries can be searched, simultaneously. Hits are listed according to quality, not according to search order.

The following are examples of proper citations to use when working with one

of the two NIST Mass Spectral Libraries and the NIST MS Search Program.

#### Search of an electron ionisation spectrum

Identification of an unidentified compound's mass spectrum was accomplished using the NIST Mass Spectral Search Program's, v.2.4, EI Normal Identity Search of the NIST 20 NIST/EPA/NIH EI Mass Spectral Library (mainlib and/or replib) and [name(s) of any other library(ies) searched]. If any Search Constraints were used, these should also be listed; especially if the Retention Index Database was used.

#### Search of a product-ion mass spectrum

Search of the NIST 20 Tandem Library (hr\_msms\_nist, lr\_msms\_nist, apci\_msms\_nist or (biopep\_msms\_nist) and [name(s) of any other product-ion mass spectral library(ies) searched] was done using the NIST MS Search Program's (v.2.4) MS/MS Hybrid Similarity Search. If any Search Constraints were used, these should also be listed.

Again, it can't be over emphasised that the library publisher, library edition and software used should be cited in detail.

This Letter to the Editor is being submitted to multiple journals in order to obtain as much coverage as possible on this topic.