

# 1998 Version of the NIST/EPA/NIH Mass Spectral Library, NIST 98

O. David Sparkman, consultant, and Doug Agnew  
Agilent Technologies Co., Palo Alto, CA

## NIST 98 Technical Note

## Agilent Technologies MSD ChemStation

### Ten Years of Focusing on Quality

It was six years ago that the National Institute of Standards and Technology (NIST) published an updated version of the NIST/EPA/NIH Mass Spectral Library. Since the release of that earlier version, NIST has completed a 10-year program undertaken to evaluate, improve, and expand the library. The results of that extensive improvement effort are now available in the NIST 98 Library.

Using a GC/MSD, NIST generated many new spectra and acquired several important collections of high-quality spectra. As a result, the NIST 98 Library now comprises 129,136 spectra of 107,886 compounds, a 75% increase over the contents of the previous version. There is an associated structure for all but 57 of the compounds, and most of the structures for those 57 are ambiguous. There are 21,250 replicate spectra and 13,205 compounds with replicate spectra.

The purpose of NIST's evaluation efforts was to reduce errors to the lowest possible level while providing the highest quality spectra possible. Even with its emphasis on quality, the NIST 98 Library favorably compares in quantity with any available library. Although the largest alternative library contains more spectra, NIST 98 contains more spectral information, providing 20% more peaks and significantly more chemical structures. The quality difference is also evident when NIST's average/median number of peaks per spectrum of 93/78 is compared with that (35/10) of the alternative library.<sup>1</sup>

Because of the large number of spectra, the possibility of duplication is significant and, indeed, duplication has been a problem in all previous mass spectral libraries. The way to ensure that duplication is not a factor is to associate a unique identifier with each unique spectrum. This is best accomplished by using a Chemical Abstracts Service (CAS) registry number or a structure. More of the NIST 98 spectra have unique identifiers than those of any other mass spectral library that has been distributed. Of the more than 107,000 compounds in the NIST 98 Library, 99.95% have an associated

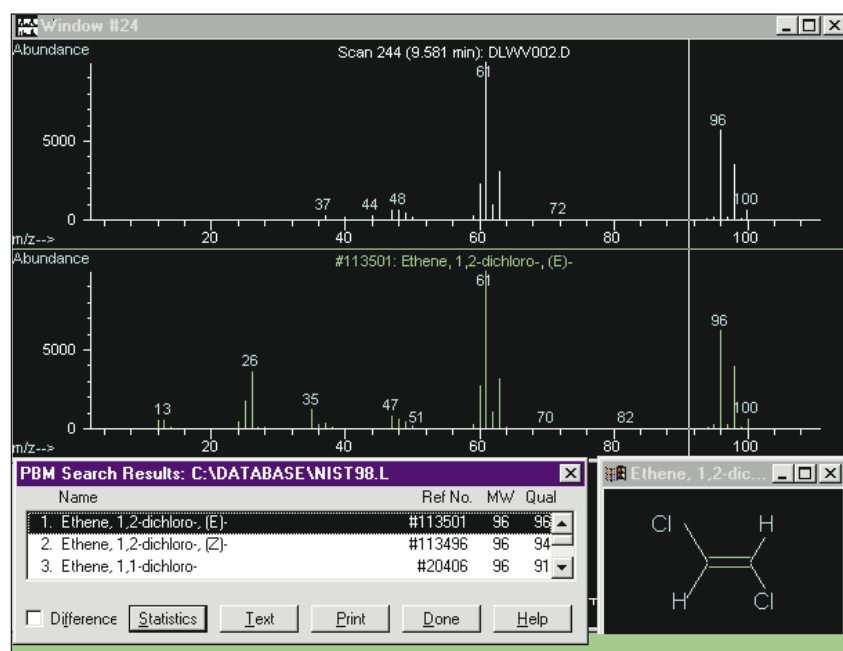


Figure 1. Example of NIST Library Search from the MSD ChemStation software.

structure. These unique structures were compared using one of the many in-house-developed software programs NIST uses to ensure the highest quality. It should be noted that there is a difference between replicate spectra (multiple spectra of the same compound from different sources) and duplicate spectra (the presence of the same spectrum in the library but with different index numbers).

Agilent Technologies distributes the NIST 98 Mass Spectral Library along with the NIST Mass Spectral Search Program for Windows. This allows the use of either the search algorithm from within MSD ChemStation or the NIST search algorithm, which has been described in numerous publications.<sup>2-4</sup>

### Re-evaluation of Spectra

In an effort to develop a Library optimized for identifying unknown compounds by their mass spectra, NIST undertook a program in which experienced mass spectrometrists evaluated complete-as-possible spectra for the presence

or absence of peaks based on a compound's structure, empirical formula, and molecular weight. When an anomaly was found, a decision (agreed upon by at least two of the spectrometrists) was made—to delete the entry from the library, to remeasure it (always done if possible), or to remove contaminant peaks. Details of the Library's development were described at several meetings of the American Society for Mass Spectrometry and may be reviewed in references 5 and 6.

The NIST 98 Mass Spectral Library now includes important collections of mass spectra from the following sources: Chemical Concepts, including Prof. Henneberg's industrial chemicals collection; Georgia and Virginia Crime Laboratories; TNO Flavors & Fragrances; AAFS Toxicology Section, Drug Library; Association of Official Racing Chemists; St. Louis University Urinary Acids; and VERIFIN and CBDCOM Chemical Weapons.

### NIST 98 Customized for the MSD ChemStation

The NIST 98 Library's function within the MSD ChemStation has been improved. In the past, all libraries used with the MSD ChemStation have required that a CAS registry number be associated with the spectrum in order for a structure to be displayed, and that required the G1036A Structures Library. Now, however, the new NIST 98 Library includes functionality that allows the MSD ChemStation software to display a structure for all spectra, whether or not there are CAS registry numbers (excluding the 57 spectra for which structures do not yet exist).

### Enhancements to the MSD ChemStation's Use of Libraries

One of the features of the new NIST MS Search Program, V1.6b, is its ability to include user-generated structures in the form of MOL files in user libraries. This feature has been a part of the MSD ChemStation, but it is

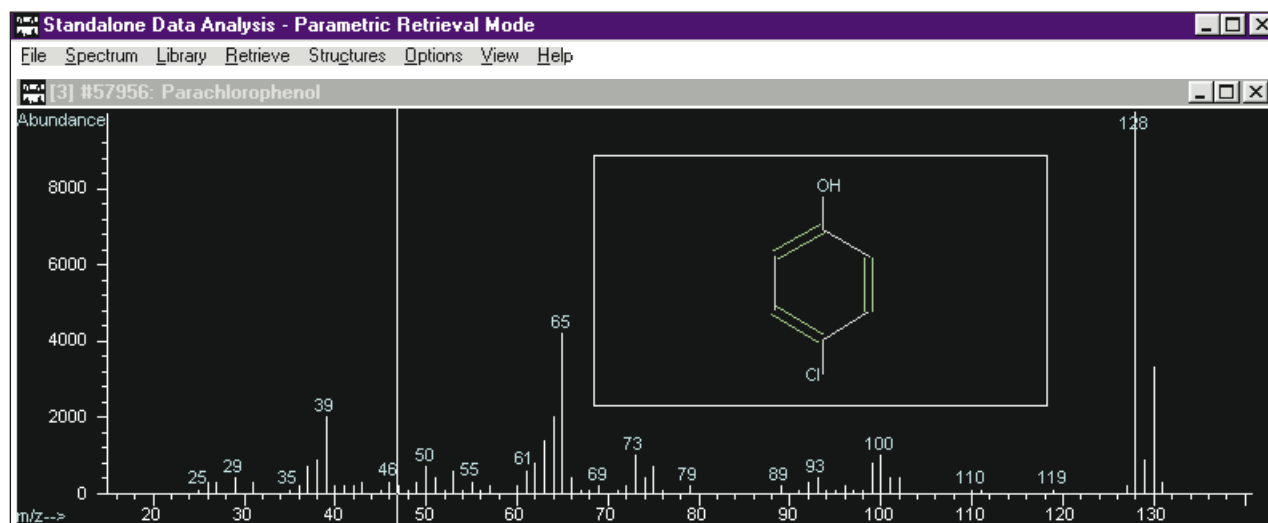


Figure 2. Importation of a MOL file structure into ChemStation User library.

now improved in the software so that implicit hydrogens associated with functional groups are displayed. See Figure 2.

## New Features in the NIST MS Search Program

There are two important new features in the NIST MS Search Program: an *Automated Mass Spectral Deconvolution and Identification System (AMDIS)*; and a unique routine that will aid in spectral evaluation and interpretation, the *MS Interpreter*.

### Automated Mass Spectral Deconvolution and Identification Software

AMDIS will read and display ChemStation GC and LC MSD data files. The files are evaluated on the basis of spectral uniqueness. Unique spectra (with contaminating peaks eliminated—deconvoluted spectra) are compared against target libraries or are sent to the NIST MS Search Program for identification. AMDIS is provided with individual target libraries (all derived from the NIST 98 Library) for use with environmental, drugs of abuse, toxicological, and flavor/fragrance applications. The libraries can be expanded and user libraries can be built from chromatographic/mass spectral data or from additional spectra from the NIST 98 or other libraries in the NIST MS Search Program format. Additional information on AMDIS can be found in Application Note (23) 5967-5735E, "Automated Mass Spectrometry Deconvolution and Identification System—AMDIS."

### MS Interpreter

Just one of the many features of the MS Interpreter is that it enhances the popular ISOFORM utility included with previous versions of the NIST MS Search Program. ISOFORM is used to calculate and display (graphically and numerically) isotopic patterns based on user-defined formulas and to produce formulas for neutral fragments and ions based on molecular formulas, elemental constraints, or  $m/z$  values of ions and neutrals.

In addition to all the functions of ISOFORM, the MS Interpreter makes possible a graphical comparison of observed and theoretical isotopic patterns, the use of a graphical tool to determine and display the  $m/z$  difference between a designated

precursor peak and another peak, and, based on a simple single-bond cleavage presumption, the display of the portion of a molecular structure represented by individual peaks in the mass spectrum. This powerful utility adds to the use of the NIST MS Search Program the NIST 98 Library in the identification of compounds whose spectra are not in the Database.<sup>7</sup> See Figure 3.

### NIST MS Search Program for GC/MS and LC/MS

The NIST MS Search Program, V1.6b, is still provided with the three search algorithms that have made it such a widely used utility in mass spectrometry laboratories: the Identity Search for spectra of compounds whose spectra are probably in the Library and the Similarity and Neutral Loss

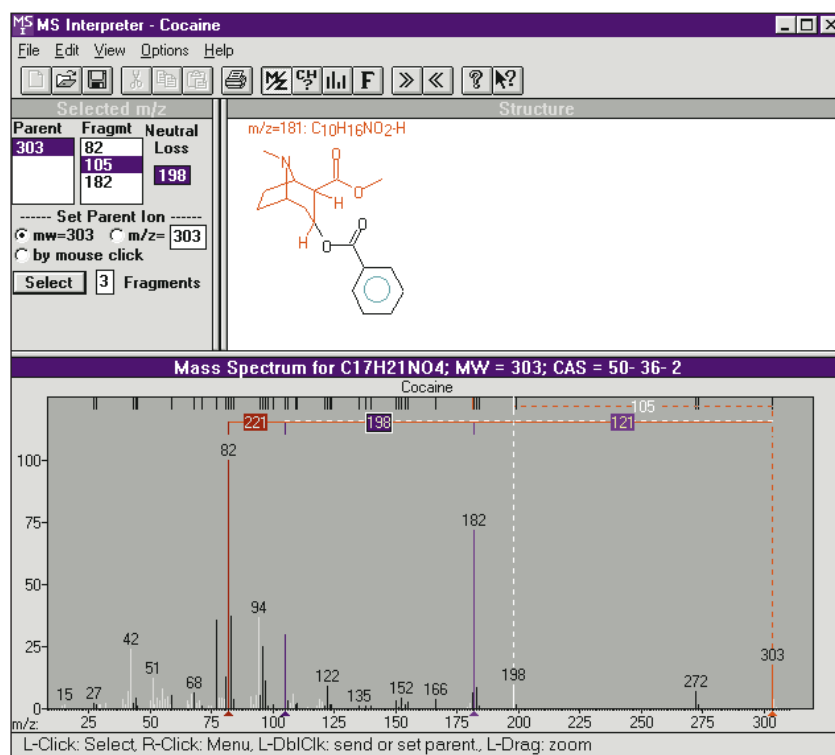


Figure 3. MS Interpreter.



Searches for spectra of compounds whose spectra are probably not in the Library. This capability, combined with Substructure Identification, is one of the reasons that the Search Program is being extensively used in the evaluation of atmospheric pressure chemical ionization (APCI) and electrospray ionization (ESI) LC/MS spectra obtained by in-source collision-induced dissociation (CID) on the LC/MSD.

The NIST Search Program allows multiple desktop configurations with up to seven independently configured Windows to examine search results and match individual needs. Spectra can be collected in the LC or GC MSD programs and automatically searched using the NIST MS Search Program. Spectra can be retrieved in the NIST MS Search Program from the hundreds of thousands of common names in the Library. As with the previous version of the NIST MS Search Program distributed, this version has the HP2NIST utility that allows for the copying of user libraries that are in the HP format to the NIST format. Additional information on the NIST MS Search Program can be found in Application Note (23) 5965-8424E, "A New Dimension in Mass Spectral Library Searching."

## Summary

Agilent Technologies' implementation of the NIST 98 Mass Spectral Library in both the NIST format, for use with the NIST MS Search Program, and in the HP format, for use with ChemStation's internal search programs, gives an unprecedented use of this powerful Mass Spectral Library.

## References

1. The figures given here are based on the reprint of a paper distributed by F. W. McLafferty et al. in conjunction with a Poster Presentation at the 45th ASMS meeting in Palm Springs, Calif., **1997**.
2. Stein, S.: Estimating Probabilities of Correct Identification from Results of Mass Spectral Library Searches. *J. Am. Soc. Mass Spectrom.*, vol. 5, **1994**, pp. 316-323.
3. Stein, S.; and Scott, D.: Optimization and Testing of Mass Spectral Library Search Algorithms for Compound Identification. *J. Am. Soc. Mass Spectrom.*, vol. 5, **1994**, pp. 859-866.
4. Stein, S.: Chemical Substructure Identification by Mass Spectral Library Searching. *J. Am. Soc. Mass Spectrom.*, vol. 6, **1995**, pp. 644-655.
5. Mikaya, A.; Stein, S.; et al. The NIST/EPA/NIH Mass Spectral Database: Simultaneous Control of Quality and Quantity. *Proceeding of the 45th ASMS Conference on Mass Spectrometry and Allied Topics*, Palm Springs, Calif., **1997**.
6. Zaikin, V.; Ausloos, P.; et al. The Evaluation of a Comprehensive MS Reference Library. *Proceeding of the 43rd ASMS Conference on Mass Spectrometry and Allied Topics*, Atlanta, GA., **1995**.
7. Robert Mistrik's Cluster Analysis research reported at the 1997 meeting of the ASMS at Palm Springs, Calif.

**Agilent Technologies shall not be liable for errors contained herein or for incidental or consequential damages in connection with the furnishing, performance or use of this material.**

**Information, descriptions and specifications in this publication are subject to change without notice.**

**Copyright © 1999  
Agilent Technologies Company  
All rights reserved. Reproduction and adaptation is prohibited.**

**Printed in the U.S.A. December 1999  
(23) 5967-5679E**