

# NIST Standard Reference Database 1A

---

## Mass Spectral Libraries (NIST/EPA/NIH EI & NIST Tandem 2020) and NIST Mass Spectral Search Program (Version 2.4)

For Use with Microsoft® Windows

### User's Guide

---

#### The NIST Mass Spectrometry Data Center

Stephen E. Stein, Project Leader

William Wallace, Group Leader

#### *Evaluation:*

Weihua Ji (EI)

James Little (EI)

Sandy Markey (EI)

W. Gary Mallard (EI)

Pedatsur Neta (Tandem)

O. David Sparkman (EI)

Edward White V (EI)

Vladimir Zaikin (EI)

Igor Zenkevich (RI)

#### *Measurement & Evaluation:*

Rabe Andriamaharavo (EI)

H. Martin Garraffo (EI)

Lorna de Leoz (Tandem)

Yuxue Liang (Tandem)

Long Pu (EI)

Connie Remoroza (Tandem)

Nino Todua (EI)

Kirill Tretyakov (EI)

Yufang Zhang (EI)

#### *Programmers:*

Yuri Mirokhin

Dmitrii Tchekhovskoi

Xiaoyu Yang

#### *Documentation and Help:*

W. Gary Mallard

O. David Sparkman

March 2020

#### U.S. Department of Commerce

National Institute of Standards and Technology

Standard Reference Data Program

Gaithersburg, MD 20899

The National Institute of Standards and Technology (NIST) uses its best efforts to deliver a high-quality copy of the Database and to verify that the data contained therein have been selected on the basis of sound scientific judgment. However, NIST makes no warranties to that effect; and NIST shall not be liable for any damage that may result from errors or omissions in the Database.

Use of any NIST Mass Spectral Database is limited to a single computer by a single user at any one time. There is no restriction as to multiple copies in the same or multiple formats on the same single-user computer or to backup copies which are for management purposes and not for use on any other computer. In the event of an upgrade from a previous version of any Database, that previous version of the Database may not be installed on another computer and used in the same time period as the upgrade. The previous version may continue to be used on the same single-user computer on which the upgrade is installed.

© 1987, 1988, 1990, 1992, 1994, 1995, 1996, 1997, 1998, 1999, 2000, 2001, 2002, 2005, 2008, 2011, 2014, 2017, 2020 copyright by the U.S. Secretary of Commerce on behalf of the United States of America. All Rights Reserved.

No part of this Database may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior written permission of the distributor. Portions of this program are also copyrighted by Microsoft Corp. and © 1984–1996 FairCom Corporation.



Microsoft is a registered trademark of Microsoft Corporation, and Windows is a trademark of Microsoft Corporation. ISIS is a trademark of MDL Information Systems, Inc. "FairCom" and "c-tree Plus" are trademarks of FairCom Corporation and are registered in the United States and other countries. All Rights Reserved.

**Certain trade names and company products are mentioned in the text to specify adequately the computer products and equipment needed to use this software. In no case does such identification imply endorsement by the National Institute of Standards and Technology of these computer products and equipment, nor does it imply that the products are necessarily the best available for the purpose.**

---

# Contents

Preface.....	1
Installation .....	3
OS Compatibilities.....	4
Features in NIST 20 Libraries and the MS Search Program .....	5
NIST 20 Libraries.....	5
Version 2.4 of the MS Search Program .....	5
Enhancements to Utility Programs Since NIST 17.....	7
Significant Enhancement to MS Search v.2.3 Released with NIST 17 .....	7
GC Methods / Retention Indices Library .....	9
The NIST Tandem Library .....	10
Software Changes from earlier versions of the Search Program .....	12
Version 2.0g of the MS Search Program.....	12
Version 2.0f of the MS Search Program.....	13
Guide to the Program .....	14
The Lib. Search Tab .....	15
Title Bar, Tool Bar, and Status Bar .....	21
Neutral Loss Display .....	23
The Other Search Tab .....	25
The Names Tab .....	27
The Compare Tab.....	28
The Librarian Tab .....	28
Spectral Analysis Utilities for NIST 20 .....	29
Automated Mass Spectrometry Deconvolution and Identification System (AMDIS) .....	29
MS Interpreter.....	30
Example Searches .....	34
Library Spectrum and Structure Searches .....	34
Library Search Options .....	37
MS/MS.....	40
Libraries .....	40
Retention Index Data .....	40
Automation.....	42
Limits .....	42
Constraints.....	42
The Search Menu .....	43
Search by Any Peaks.....	44
Exact Mass Search.....	47
Search by Name - Browse .....	48
Sequential Method.....	48
Basis for Interpretation of the Library Search Results .....	49
Substructure Information .....	50

---

Use with Instrument Data Systems .....	54
Use with Third-Party Drawing Programs .....	57
User Libraries .....	58
NIST Text Format of Individual Spectra.....	61
Import and Export of Retention Indices with a Spectrum.....	62
RI Export with a Spectrum in a MSP or JCAMP file.....	62
RI Import with a Spectrum in a MSP or JCAMP file .....	63
APPENDIX 1: Creating AUTOIMP Files Used in NIST MS Search .....	64
APPENDIX 2: Using the NIST MS Search Program with ChemStation.....	65
APPENDIX 3: Copy ChemStation and Other Database to NIST User Library Format .....	68
APPENDIX 4: Using the NIST MS Search Program with Thermo Fisher Corporation Xcalibur Software .....	71
APPENDIX 5: Search Algorithms .....	72
Screening.....	72
Search .....	73
APPENDIX 6: Syntax for Some Constraints in Searches.....	74
Syntax of Name Fragments Constraint.....	74
Syntax of Tags in Comment Constraint .....	74
Tags in Comment Constraint for MS/MS Text Information.....	79
MS/MS Instrument Type Constraint.....	80
APPENDIX 7: Contributing EI Mass Spectra to the NIST/EPA/NIH Mass Spectral Library .....	81
APPENDIX 8: Support Contacts.....	83

---

## Preface

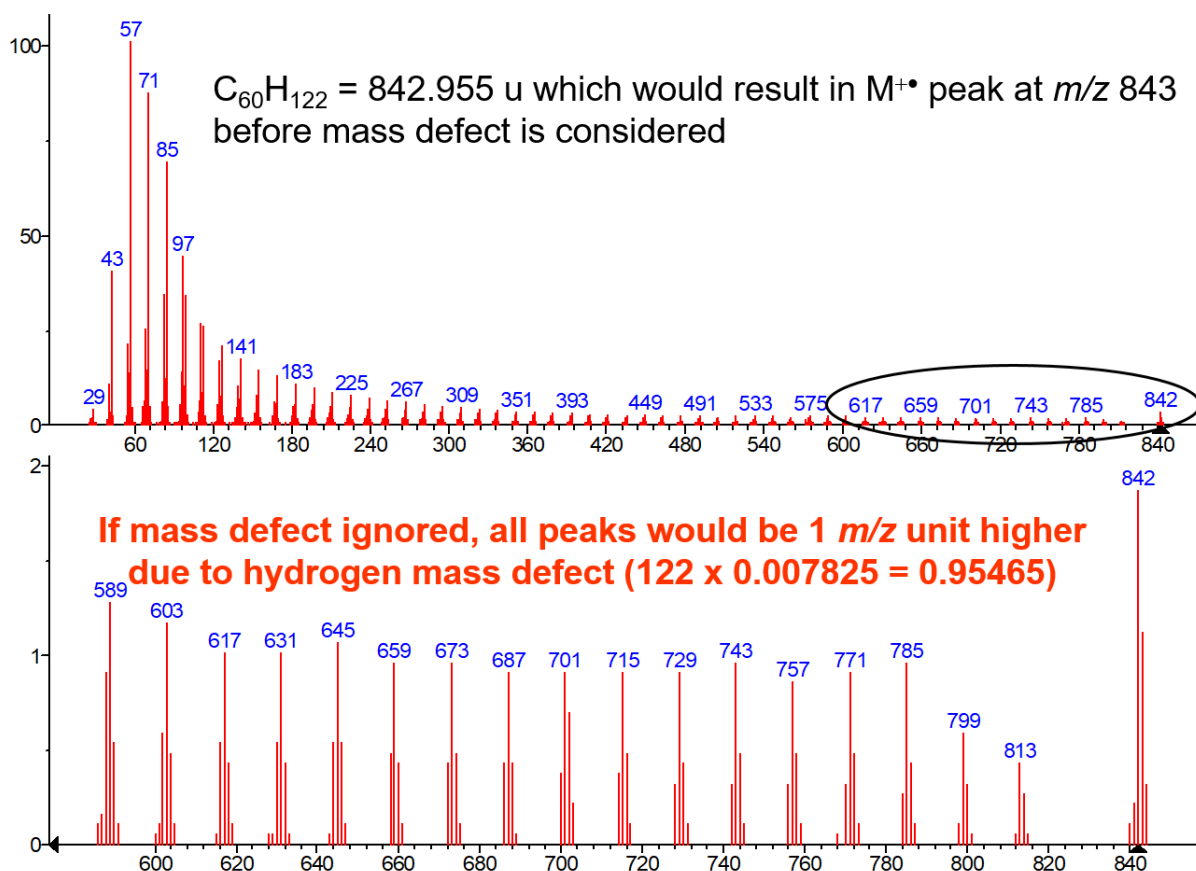
Even before the introduction of the first commercial mass spectrometers in the very early 1940's, individuals and organization began collecting reference mass spectra. These spectra displayed (first tabular and later graphic) the  $m/z$  values as integers. This was partly because of the limited accuracy of the instruments used to measure the mass of organic compounds at that time. Even though the mass spectrometer measures the actual  $m/z$  value of ions, the upper limit was such that cumulative effects on the mass of an ion were inconsequential. The mass of all the elements are measured relative to the mass of the most abundant atom of carbon ( $^{12}\text{C}$ ) having been assigned an exact value of 12 Da. Mass spectrometers do not measure the mass of ions; they measure the mass to charge ratio ( $m/z$ ) values of the ions. This is the mass of the ion divided by the number of charges that the ion has. In the case of electron ionization (EI) nearly all ions have only a single charge; therefore, because the charge on the positively charge single-charge molecular ion ( $\text{M}^{+}$ ), formed by a molecule's loss of a single electron, which is  $\sim 1/2000$  of the mass of either one of an atom's neutron or proton, it is safe to say, on an integer scale, that the mass of the molecule and the  $m/z$  value of its  $\text{M}^{+}$  are the same.

The EI mass spectrum of an organic compound is represented by a series of peaks sometimes starting with a cluster of peaks representing its  $\text{M}^{+}$  peak and preceded to lower  $m/z$  values, the lowest of which is usually a function of the  $m/z$  acquisition range. These peaks represent ion formed by the fragmentation of the  $\text{M}^{+}$ . Examination of the  $\text{M}^{+}$  peak sometimes results in the observation of one or more peaks at higher  $m/z$  values usually one  $m/z$  unit apart. These peaks represent ions of the same elemental composition as the  $\text{M}^{+}$ , but of different isotopic compositions. These are isotope peaks. EI mass spectrometry of organic compounds is generally limited to 11 elements (C, H, N, O, Si, S, P, and the halogens). Some of these elements have multiple isotopes (atoms of a specific element that have different numbers of neutrons, whereas the number of protons and electrons are the same). Others are what is known as monoisotopic elements, meaning that they do not have isotopes. These are F, P, and I.

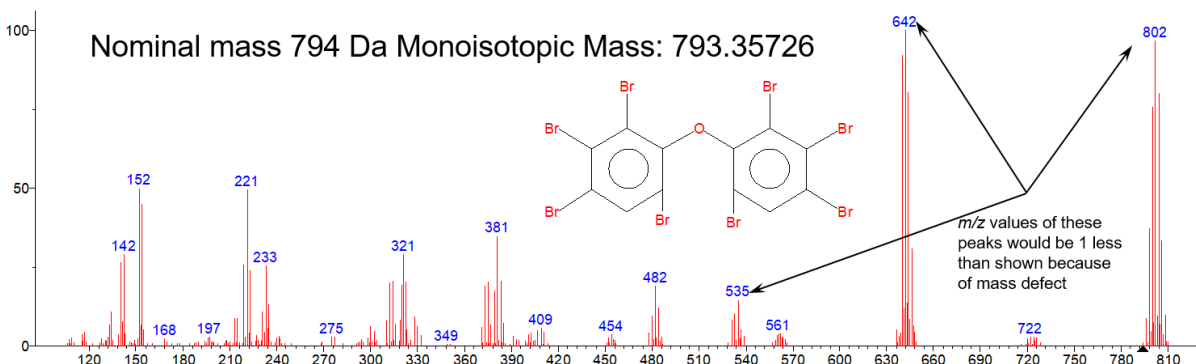
The lowest  $m/z$  value peak in the cluster of peaks representing an ion of a specific elemental composition and is called the **monoisotopic peak**. This peak represents an ion that has all the atoms of each element being that element's most abundant isotope. The exact mass of an element's most abundant isotope is that elements **monoisotopic mass**. The integer value for the most abundant isotope of an element is called its **nominal mass**. The **nominal mass** of an ion is the sum of the **nominal masses** of each of its elements times the number of atoms of that element.

The difference between the integer mass of a nuclide and its exact mass is called the **mass defect**. The **mass defect** of  $^1\text{H}$ ,  $^2\text{H}$ ,  $^{13}\text{C}$ , and the two isotopes of nitrogen is positive; meaning that the exact mass is a little greater than the integer mass. The mass defect of the nuclides of all the other elements listed above is negative; the exact mass is a little less than the integer mass. The **mass defect** is always less than 100 milidalton regardless of it being positive or negative. The two elements with the greatest **mass defect** relative to their exact masses are hydrogen (7.825 Da relative to an exact mass of 1.007825, 0.8%) and bromine (21.7 Da relative an exact mass of 78.9183, 0.03%).

As the mass range of instruments increased and the accuracy improve a decision had to be made as what would be the standard for ions in reference libraries. Would it be based on **nominal mass** or the mass measured by the mass spectrometer. Consider the mass spectrum of Hexacontane ( $\text{C}_{60}\text{H}_{122}$ ). The measured integer  $m/z$  value for the  $\text{M}^{+}$  would 843; whereas the **nominal mass** of the  $\text{M}^{+}$  is 842 Da. It was decided that all refence spectra would have peaks which corresponded to the ions **nominal mass**. Some instrument companies' data systems will adjust the EI to accommodate the mass defect. If compensation is not done the acquiring instrument, then when the spectrum is imported into MS Search, MS Search will make the correction. All spectra in the NIST/EPA/NIH EI Reference Library are based on **nominal mass**. There are spectra in the NIST Tandem Library that are based on measured mass and these are what are called accurate  $m/z$  value spectra, sometime called high-resolution spectra.



NIST/EPA/NIH NIST 20 Spectrum of Hexacotane ( $C_{60}H_{122}$ ).



NIST/EPA/NIH NIST 20 Spectrum of 2,2',3,3',4,4',6,6'-Octabromodiphenyl ether.

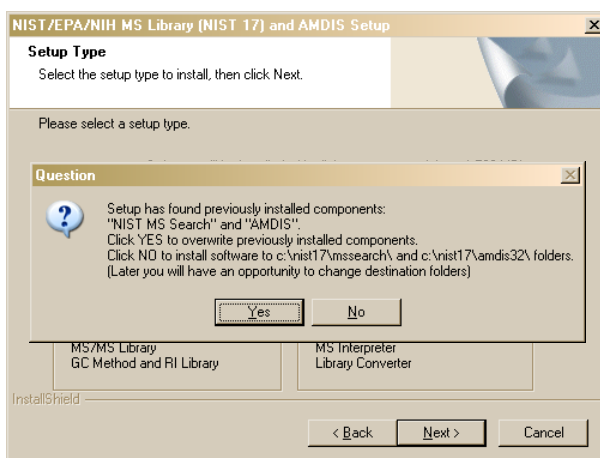
# Installation

## Windows XP SP3, Windows Vista, Windows 7, 8, 8.1, 10

- Open Windows Explorer and navigate to the location of the NIST 20 installation. This may be a folder, on a CD drive, or on a USB drive. Run setup.exe from the CD.
- In the folder or on the drive which contains the NIST 20 install-software, highlight SETUP.EXE and double-click the left Mouse button.
- Follow the instructions presented by the SETUP program.  
In upgrading from a previous version of the NIST MS Library, there is a possibility to keep both versions.<sup>1</sup> The Installer may behave differently under various versions of Windows due to User Account Control security features.

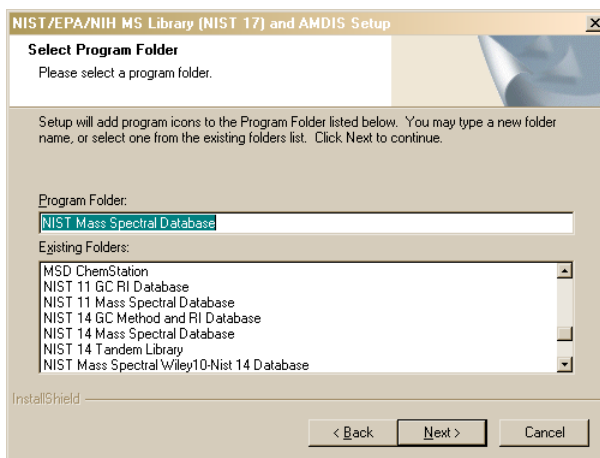
<sup>1</sup> In upgrading from a previous version, the Libraries and Programs will be overwritten by default. To avoid this, (1) the Libraries should be installed in a new folder, and (2) a new Program Folder should be created (selected) for the Start Programs shortcuts.

(1) To retain the previous version of the **NIST MS Search Program** and its associated Libraries, select **No** in the **Question** dialog box below.



**Question dialog box: Select overwriting the previous installation, or retain the previous installation and install NIST 20 in a new folder.**

- (2) When the dialog box shown below appears, the name of the Program Folder must be changed to **NIST 20 Mass Spectral Database**. If this is not done, the contents of the existing Program Folder, NIST Mass Spectral Database, will be overwritten, and the previous version of **MS Search** and its Database will no longer be accessible from the **Start** menu.



**The name of the Program Folder must be changed if a previous version of MS Search and its Database are to be retained.**

---

## OS Compatibilities

**NIST 20** and **MS Search** are **NOT** compatible with Windows 3.x, 95, 98, Me, NT 3.x, Windows NT 4.0, Windows 2000, and Windows XP before SP3.

**NIST 20** and **MS Search** are fully compatible with Windows XP SP3, Vista, Windows 7 (XP, Vista, and Windows 7 in both 32- and 64-bit editions), Windows 8, and Windows 10\*.

Also needed are:

- **Adobe PDF** reader to access the **NIST 20** Manual
- WinNT4Sp3+MSIE4.0/4.01/4.01SP1+hhupd.exe or WinNT4Sp6+MSIE4.01SP2 or later version of IE to access **NIST MS Search** html Help

Windows NT 4.0 SP1 and WinNT4Sp3+MSIE4.01Sp1+hhupd.exe under MS Virtual PC 2007 are the earliest versions of Windows that **NIST 20** and **MS Search** have been tested with.

\* Issues with the *custom scaling level* feature of Windows 10 may have been observed. It is best to turn the *custom scaling level* feature in Windows 10 off.



---

# Features in NIST 20 Libraries and the MS Search Program

## NIST 20 Libraries

The 2020 version of the **NIST Mass Spectral Libraries (NIST 20)** contains seven mass spectral libraries and a library of GC retention indices and methods.

Contents of the NIST/EPA/NIH (NIST 20) Mass Spectral Libraries		
Library Name	Library Description	Library Contents
mainlib	Main EI MS Library	306,643 Compounds
replib	Replicate spectra Library	43,774 Alternate spectra of compounds in mainlib
hr_msms_nist	Accurate-mass Tandem Library of small molecules	1,026,712 Accurate-mass Tandem spectra for 27,840 compounds
lr_msms_nist	Low-resolution Tandem Library of small molecules	215,649 Low-resolution Tandem spectra for 28,559 compounds
apci_msms_nist	Library for 'Extractables and Leachables'	3,555 Accurate-mass Tandem spectra for 246 compounds
biopep_msms_nist	Tandem Library of biological peptides	90,244 Tandem spectra for 1,904 peptides
nist_ri	Retention Index/GC Method Library	447,289 Values for 139,382 compounds (114,629 in mainlib)

The first two contain EI spectra, the next four have tandem (MS/MS) spectra and the last contains gas-phase retention indices. Since the 2017 release, these libraries have grown in coverage by the fastest rates since their inception. The NIST/EPA/NIH EI Library has increased by nearly 40,000 Compounds. The Retention Index library showed a corresponding increase. The Tandem library had doubled in coverage, now including nearly 31,000 compounds. This includes 1.3 M spectra (low and high resolution) of 186 K precursor ions.

Also, for the first time, most new spectra were measured at NIST based on their analytical importance and then evaluated by at least two evaluators with the aid of newly developed software.

There is increased display of chirality in the structures. Over the last several releases of the NIST Libraries, more and more structures were displayed with chirality.

## Version 2.4 of the MS Search Program

Since the release of Version 2.3 with NIST 17 no substantial changes have been made to the fundamental Search Routines. There have been a number of display changes which allow for better control of the display of Hit Lists and the handling the many Tandem spectra that can be associated with a single compound.

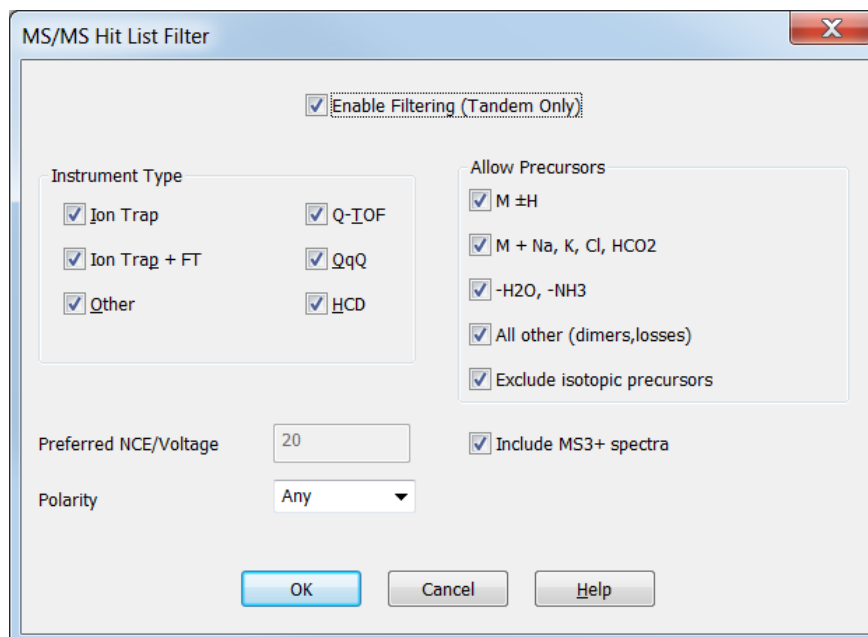
- Names and orders of some library searches have been changed

Identity Search		Similarity Search	
NIST 20	NIST 17	NIST 20	NIST 17
EI Normal	Quick	EI Hybrid	Hybrid
EI Quick	Normal	EI Simple	Simple
MS/MS	MS/MS	EI Neutral Loss	Neutral Loss
In-Source Hi Res	In-source Hi-Res	MS/MS in EI	MS/MS in EI
		MS/MS Hybrid	MS/MS Hybrid

EI searches are strictly unit mass resolution and MS/MS (Tandem) searches allow the specification of high mass accuracy peaks.

- Tandem Mass Spectra Filtering for all Hit Lists (Spectrum and Other Searches)*

The NIST Tandem Library contains spectra over a range of energies, precursor types and fragmentation methods. This means that a search of the Tandem Libraries can produce many Hits generated from a single compound, which can be tedious to sort through. To facilitate this task, a filtering process is available that can easily control the types of spectra shown in a hit list. This is invoked by pressing the filter button in the button bar above the hit list showing results. This is available for both the Lib Search and Other Search tab views. The filtering process may be turned on and off with the control at the top of this dialog box (below). When the filter is enabled, the filter button will appear to be pressed. Upon pressing OK, the hit list will reflect the selections in this dialog box. These control match those in the Constraints section, except they do not permanently removed hits – all original hits can be seen by turning off the 'Enable Filtering' checkbox. This filter has no effect on EI searches.



**Figure 1: Tandem Hit List Filter Used to Limit Hits Displayed from Library and Other Searches Only Applicable to the Tandem Libraries**

- 
- Corresponding changes have been made in the Constraints used when searching Tandem Libraries.

## ***Enhancements to Utility Programs Since NIST 17***

**Mass Spectral Interpreter** has significantly evolved over the past three years. There was a major release in April of 2020 that include a considerable number of enhancements from earlier versions. The major areas involve mechanisms and their relative rates for fragmentation of all types of ions (EI and Tandem MS, both positive and negative; integer and accurate mass) and well as various user interface enhancements. This program connects mass spectral peaks to their probable chemical structure origin and can be quickly invoked from the NIST MS Search Program with the F9 key.

## ***Significant Enhancement to MS Search v.2.3 Released with NIST 17***

Following Version 2.2, distributed with NIST 14, there were a number of important features added to the NIST Mass Spectral Search Program. These are summarized below.

### Features Relevant to Electron Ionization Mass Spectra (Low Resolution)

- A new EI Hybrid Search (Similarity/Hybrid)
- Presearch for EI Similarity/Neutral Loss search

### Features Relevant to All Mass Spectra

- InChIKey presearch for all Library spectrum searches (finds all spectra, which have the same first InChIKey segment as the search spectrum, and compares them to the search spectrum)
- When a spectrum is associated with a structure in the Spectrum Information dialog box of the Librarian tab view, the InChI Key is automatically associate with the spectrum.
- Columns in any Hit List can be positioned as desired.
- Update list of libraries may be performed without exiting MS Search
- Indexing old libraries for EI Hybrid, EI Loss, HiRes MS/MS Hybrid, HyRes Any-peaks Peak and Loss searches, and InChI presearch. This includes third party libraries like the Wiley Registry.

### Features Relevant to Tandem and High-Resolution Mass Spectra

- HiRes MS/MS Hybrid Search
- Types of Any Peak search: HiRes peak and HiRes loss
- Display additional MS/MS information in hit lists
- Edit accurate peak  $m/z$ , intensities and peak annotations in the built-in mass spectrum editor
- MS/MS Precursor  $m/z$  search
- Precursor type constraint
- Ion polarity filter

---

## The following features were added to version 2.2, distributed with NIST 14 as compared to ver. 2.0g, distributed with NIST 11 or with NIST 12 Tandem Library

### Features Relevant to Electron Ionization Mass Spectra (Low Resolution)

- Use of Retention Index in library searching: Deviation of experimental RI from the library value optionally used to penalize the Match Factor.
- Older user libraries with RI values can be indexed to allow the RI values to be used in library searching
- Display RI values in a hit list column
- Import, export, and edit compound RI value(s)
- New definitions of derivatives are used for finding replicate spectra
- Optionally display derivative precursor in the Lib. Search hit list

### Features Relevant to All Mass Spectra

- All small molecule mass spectra in NIST libraries have InChIKeys linked to PubChem (NIH). These InChIKeys may also be used for searching compound information on the Web or for exact structure searches within NIST libraries.
- Set Prepend/Overwrite/Ask automatic mass spectrum import option.
- Import and display Glycan structures in KEGG Chemical Function (KCF) format and add them to user library mass spectra

### Features Relevant to Tandem and High-Resolution Mass Spectra

- Display additional MS/MS information in hit lists
- Edit accurate peak m/z, intensities and peak annotations in the built-in mass spectrum editor
- In-source tandem spectra or EI with accurate m/z spectra - High resolution mass spectra
  - Import of 'in-source/EI with accurate m/z' spectra, which do not include a defined m/z of the precursor ion. To import these spectra, select "In-source /EI (accurate m/z)" in Spectrum Import Options.
  - In-source/EI with accurate m/z spectra have accurate ion peak m/z and intensities. The accuracy is set in Spectrum Import Options.
  - In-source/EI with accurate m/z spectra may be searched in In-source/EI with accurate m/z and/or MS/MS library with the new In-source HiRes search (ion peak m/z tolerance down to 0.015 ppm or  $6 \times 10^{-5}$  m/z units), Similarity Simple, Identity Normal or MS/MS Presearch OFF search options. For this purpose, adding Reverse Search option may be useful.
  - In-source/EI with accurate m/z spectra may be added to a user library. Currently, in-source/EI with accurate m/z spectra added with NIST MS Search to a user library may be searched with In-source HiRes search only with Presearch OFF option. Rebuild the library with Lib2NIST to get in-source HiRes search benefits<sup>2</sup>.
- A new MS/MS search for small molecules
  - Improved ability to identify compounds with a few dominant peaks – a common situation for Tandem MS.
  - More robust scoring for spectra acquired on imperfectly tuned instruments

---

<sup>2</sup>Library searchable with In-source HiRes search has files peak\_em0.inu and peak\_em0.dbu

- 
- New Mass Spectrum Import Options
    - Control the number of decimal places in the input m/z
    - Set peak intensity threshold
    - Import in-source or EI spectrum with accurate ion peak m/z and intensities.
    - Import MS/MS and in-source/EI with accurate m/z spectra from JCAMP files (October 2012, nistms32.dll version 2.1.5.3 or later)

## GC Methods / Retention Indices Library

In NIST 17, the GC Methods/Retention Indices library was extended by including summary statistical analysis of all data in the record. This information will be displayed in the spectrum text information even if the **Retention Indices** box is unchecked in the properties box. The summary data are given in terms of the type of column (standard non-polar (DB-1), semi-standard non-polar (DB-5), or polar (carbowax)), the **median** value of the retention data for that column type, the deviation and the number of experimental determinations. Note the deviation is calculated as the median absolute deviation. In the example below (for 6,6-Dimethyl-2-methylenebicyclo[3.1.1]heptan-3-ol or pinocarveol), there are 13 values for the DB5 series of data, 9 for the DB 1 and 12 for the polar column with median absolute deviations of 4,7 and 8 respectively.

Experimental RI median±deviation (#data)

Semi-standard non-polar: 1138±4 (13)

Standard non-polar: 1135±7 (9)

Polar: 1654±8 (12)

The data will also optionally be shown in the Library Search Hit List and the Hit Lists of the Searches in the Other Search tab of MS Search. The column position for the RI values is user selectable.

The Retention Indices checkbox controls the display of the detailed experimental data. The number of GC Methods/Retention Indices displayed is controlled by selecting **All** or **First** and entering the number of values to be displayed. The default is to display the first 2 values. Displaying all values will cause the display to be slow if there are a very large number of literature values.

In addition, for most compounds, values are shown using two theoretical methods:

1) **Estimated retention index** is displayed based on the structure associated with the spectrum. This is based on a chemical group additive approach. These values are reported in Kováts index units (**iu**) with a confidence interval; e.g., "Confidence interval (Ketones): 57(50%) 246(95%) iu" reported for acetone. This means that there is a 50% probability the true value falls within 57 iu of the estimated value and 95% of the time the true value falls within 246 iu of the estimated value. The analysis is based on error measurements of all the data for the class of the compound given; i.e., in this example, ketones. These estimates only shown if **Retention Indices** has been checked in the properties box. The estimates are not displayed in the Hit Lists. If spectra in a user library have associated structures, an Estimated Kováts RI may also be listed for these spectra if the calculation is possible.

2) **AI predicted non-polar retention index** values from a newly developed approach by NIST scientist Lewis Geer are also shown. These predicted non-polar Kováts retention indices [1] are more accurate than the Estimated non-polar values provided for NIST/EPA/NIH EI Library spectra in this and previous releases as well as user library spectra with associated structures. AI estimated values are not available for user libraries with structures. The AI values were calculated for semi-standard non-polar (DB-5) columns and are expressed in index units (iu). These indices are generated by feeding the chemical structure of compounds into a path-augmented graph transformer network [2]. This network was trained

on experimental retention indices from the NIST EI Library, minus a test and validation set. The test set was used to estimate 50% and 90% confidence intervals for all compounds that have predictions.

References: 1. Kováts, E.: Gas-chromatographische Charakterisierung organischer Verbindungen. Teil 1: Retentionsindices aliphatischer Halogenide, Alkohole, Aldehyde und Ketone. Helvetica Chimica Acta 41(7), 1915(1932 (1958). doi:10.1002/hlca.19580410703  
2. Chen, B., Barzilay, R., Jaakkola, T.: Path-Augmented Graph Transformer Network (2019). <https://arxiv.org/abs/1905.12712>

## ***The NIST Tandem Library***

Tandem spectra, along with an increased number of fields relevant to data obtained using MS/MS, are provided in four separate libraries (high res, low res, APCI and biological peptides, see accompanying (see also “Tandem-library-users” documentation for details) . Most have been measured at NIST, though some have been acquired arisen from other sources. The preparation of the Tandem Library revealed that at sufficiently high signal-to-noise measurement conditions, modern instruments can provide very reproducible “library-searchable” spectra. Although collision energy is an important variable in collision cell instruments, spectra vary in an understandable way depending on the compound and instrument type as well as collision-energy conditions. Energy-dependent variation is accounted for by including spectra for most compounds over a wide range of collision energies, ranging from slight to complete dissociation of the precursor ion. The score reported for the best matching spectrum is used for identification (use the ‘Best matching only’ button to show only closest matching spectra).

The **MSMS** tab, in previous releases, has been discontinued. The tandem mass spectra are treated like spectra found in any NIST MS Search Program compatible Library. Just select the NIST library or libraries you want to search for CAS#, Formula, molecular weight range, etc. All that is necessary is to include the appropriate library or libraries in the **Include Libs** column of the **Search** dialog box for the **Other Searches** or as the single library to be searched in the **Incremental Name Search**. The **Library Spectrum Search** and **Structure Similarity Search** are also applicable. One of the challenges with looking at the tandem ion mass spectra for a given compound is that there can be several different precursor ions for that compound and spectra for sever different collision energies. Some filtering can be done with the Tandem Filter feature that has been added. This filter is not available when searching by name.

It is important that the user must understand that tandem spectra in general are different from EI spectra as well as the fact that structures in the MS/MS Library refer to the form of the precursor ions prior to ionization. It should also be noted that the integers in the MW field of the Text Information window for the spectrum is the nominal (integral) mass of the neutral molecule. The Exact Mass field shows the monoisotopic mass.

For beam-type instruments (e.g., triple quadrupoles, HCD, and Q-tof instruments), the **MS2** designation in the **Spectrum Type** field refers to spectra generated by dissociation in a collision cell of a selected precursor ion. For ion trap instruments, the **MS2** designation refers to dissociation in the trap after following the isolation of the precursor ion.

In the case of data acquired using an ion trap mass spectrometer, may also be data designated as **MS3**, and **MS4** spectra, which were obtained by the isolation of a tandem as a precursor-ion for a subsequent iteration of MS/MS.

As can be seen from **Error! Reference source not found.** there is a great deal more information provided in a text format for the spectra in the tandem libraries.

The **Library Spectrum Search** and **Structure Similarity Search** are also applicable. Note that in all cases, the structure shown is that of the compound prior to ionization.

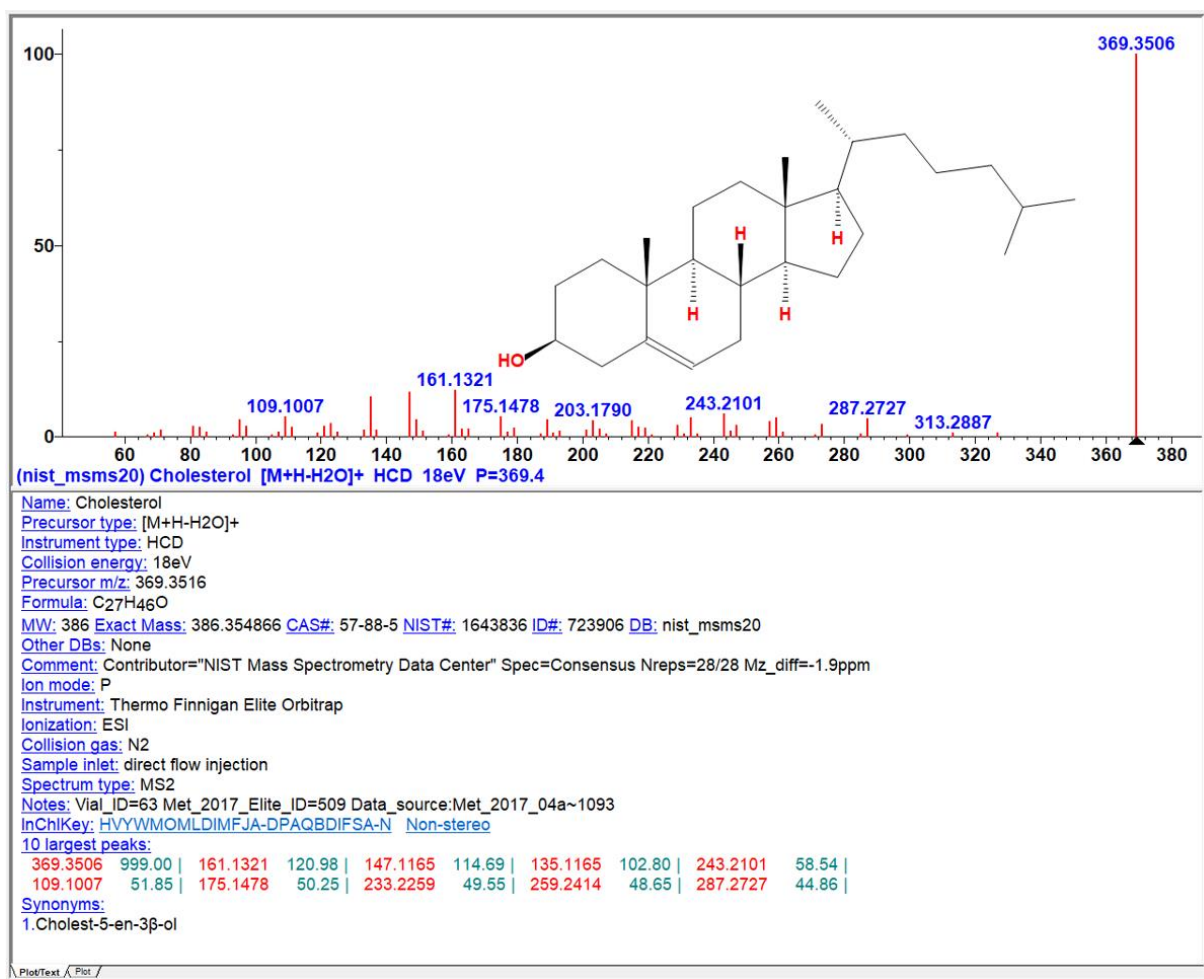


Figure 2. Mass Spectrum from the NIST 20 Tandem Library.

---

## Software Changes from earlier versions of the Search Program

### Version 2.0g of the MS Search Program

Below is a list of features new to the previous version, 2.0g, which was distributed with NIST 11.

- A new **Exact Mass Search** allows searching for exact mass of a precursor (exact molecular weight) using its mass or exact  $m/z$  of a product ion. It considers the presence or absence of an electron, adducts or losses; accepts uncertainty in ppm or millimass units; and allows searching for isotopic or monoisotopic peaks.
- A new **Exact Mass** Constraint similar to **Exact Mass Search**.
- Ability to use accurate mass to search exact fragment ion mass values in NIST libraries using the **Any Peaks Search**.
- Display of the exact mass with any spectrum that has an associated chemical formula. This even works for low resolution EI peaks – these are processes using the separate MS Interpreter formula to assign chemical formula to most peaks.
- Manage and search as many as 127 separate MS libraries instead of 16, which was the limit in previous versions.
- Optional display of columns in **Hit Lists** containing the number of synonyms and the number of Other Databases in which the Hit appears. The **Hit List** can be sorted according to the values in these columns.
- Full compatibility with most recent NIST Peptide MS/MS Libraries.
- Copy selected lines from **Hit List** or **Spec List** into Windows Clipboard as tab-delimited Unicode text; this may be conveniently pasted into Excel.
- Spectra in the MSP text format can be imported from the Windows Clipboard into the **Spec List** via a Right Mouse button (RMB) menu option.
- Searching for words in mass spectrum **Text Information** Window.
- Optional exclusion of Homologues from the results of **Structure Similarity Search**.
- Import spectra from mzXML and mzData files.
- Multiple CAS Registry numbers for a spectrum. A small fraction of spectra in NIST 14 has more than one CASrn. Each CASrn may be used in a **CASrn Search**.
- The MS Interpreter program now enables high mass accuracy for both molecular ions and product ions.
- An alternative peak matching method using ‘time warping’ was implemented to better match peaks in noisy MS/MS spectra.
- Several convenience features.<sup>3</sup>

---

<sup>3</sup> New convenience features added to version 2.0g included:

- Default view of **Constraints** and **Other Searches** dialog boxes changed to provide more convenience to the users.
- Ability to Sort the **Spec List** alphabetically, by number of synonyms, by number of other databases.
- Toggle between Tab views (**Lib. Search**, **Other Search**, **Name**, **Compare**, **Librarian**, and **MSMS** Windows) using Ctrl-Tab or Ctrl-Shift-Tab.
- When a Tag from the “Comment” field is displayed separately, its display in the “Comment” field is suppressed.



---

## Version 2.0f of the MS Search Program

Below is a list of features new to the previous version, 2.0f, which was distributed with NIST 08.

- A new spectrum search for MS/MS spectra in MS/MS libraries, including the NIST MS/MS libraries.
- Spectra of stereo isomers, tautomers, and derivatives are displayed as replicate spectra.
- Tags (which can act as Field Names) can be entered into the "Comments" field of a user library along with the Field's contents.
- **Hit List** can be sorted alphabetically.
- Five new Constraints have been added: **Tags in Comment**, **Peptide Sequences**, **Peptide Mobile Protons**, **Peptide Charge**, and **Peptide Number of Residues**. **Tags in Comment** Constraint, for example, allows searching for the name of the contributor in the EI Library.
- A search for information in the "Contributor/Comment" field, which may be particularly useful for certain types of information relevant to MS/MS.
- Flexibility has been added to searches using a "Name Fragment in Constraints". It is now possible to specify the presence or absence of any arbitrary part or parts of a name.
- The **Program** does not use or require the installation of special NIST fonts.

---

## Guide to the Program

The Program uses tab displays. The tabs:

**Lib. Search**, **Other Search**, **Names**, **Compare**, and **Librarian**, tabs

which are located at the bottom of the **Main** Window. Selecting a tab results in the display of a different Window. All tab displays have moveable divider bars (splitters), some of which can be re-orientated as well as tabs associated with various windows within a view. Selecting Window / Tile Vertically or Window / Tile Horizontally from the main menu displays all windows.



**Lib. Search** tab (Figure 1) is used for all searches based on comparison of spectra or structures to a library.

**Other Search** tab is used for most non-spectral searches other than a specific Name Search, i.e. Molecular Weight (Nominal Mass), Exact Mass, MS/MS Precursor  $m/z$ , Formula (Elemental Composition), Any Peaks, CAS number, NIST Number, and ID Number, which can be Constrained. There is a wild card (Sequential) search which allows users to create their own search criteria. These are individually described later in this manual.

**Names (Incremental Name Search)** tab allows you to type in a name and have the program match the name as you type. This search uses all names and synonyms in the specified library

**Compare** tab allows for easy comparison between spectra. It is, by default, automatically refreshed with the latest spectral search of the library. You can put your own spectra into also.

**Librarian** tab keeps a history of the recent spectral searches and also can be used for import, editing and user library construction. In some cases the Librarian tab is also referred to as the Spec. List. The upper left window in the **Lib Search** tab is also called the **Spec List**.

In every tab view, properties of the individual windows may be controlled using the **Right Mouse Button (RMB)** to display a menu from which the **Properties** dialog box may be accessed. Alternate layouts can be selected using the **Change layout** button  in the **Tool Bar** which is just under the **Main Menu**. A Tool Tip (the name of a item) is displayed when the Mouse pointer is placed over a **Tool Bar** button. There is also a **Tool Bar** button to restore the default layout  (**Set default layout**). Alternatively, Default Layout and Change Layout in the Window menu may be used.

All the tab displays have the same Main Menu. The various menu selections available for each of these tabs are described in more detail throughout the manual. The **File** menu has **Save Configuration** and **Restore Configuration** selections which allows all the various tab display configurations to be saved to a specific file name and then recalled. Multiple users may have different ways of configuring each tab that is best for them. This way, each one can restore a specific set of display options. This does not include data search history. There is no change in the search history or **Spec List** resulting from a **Restore Configuration**.

The **Search** selection on the Main Menu allows for quick selection of non-spectral type search. Depending on the choice the selection will transfer you to the correct tab and bring up the options for that search.

The **View** option on the **Main Menu** (top of Figure 1) allows the display of the **Tool Bar** and/or the **Status Bar** (bottom portion of Figure 1 containing **For Help press F1** and **Ident**). There are also options to turn off the display of the **Compare Pane** window in the **Lib. Search** and **Compare** tab views.

Tab displays can be brought into focus by placing the Mouse pointer on the tab name at the bottom of the desktop and clicking the left Mouse button. It is also possible to switch between tab views using the hot keys **Ctrl-Tab** and **Ctrl-Shift-Tab**.

Many of the controls for individual window are similar from tab to tab. These are covered in detail for the **Library Search** Tab but apply for most of the tabs.

---

## The Lib. Search Tab

The first tab (lower left of the Program's display), **Lib. Search**, is used with a search of a spectrum or structure against mass spectral libraries: a user library and/or the **NIST 20 Libraries** (Figure 1). The spectrum and structure searches can be done against multiple libraries (maximum of 127), and the selected libraries can be different from those selected for the various searches found in the **Other Search** tab. Spectra can be imported into the **Spec List** from linked mass spectral data systems software, text files containing one or more spectra in the NIST format, or from the Windows Clipboard. Structures in a MOL file format can be pasted into the **Spec List** for searching from the Windows Clipboard or an ACSII file.

The top window on the left side of the **Lib. Search** tab is **Spec List** (known as the **Clipboard Window** in previous versions of the Program). The middle window is a **Hit Histogram**, showing the distribution of Hits according to the Match Factor values. This Histogram can be displayed with either a linear or log scale on the y axis. The lower left window contains the **Hit List** at the bottom. It is important to note that depending on the user selections, in addition to the score and chemical name for each Hit, a number of other values may be displayed using settings available in the **Properties** dialog (RMB menu).

On the right side of the movable vertical divider bar are three separate windows: **Search Spectrum/Spec List Spectrum**, top; **Compare**, middle; and **Hit List Spectrum**, bottom. There are vertical divider bars associated with **Spectrum/Text** tabs of the top and bottom windows. By selecting **View** from the **Main Menu Bar**, the display of the **Compare** window can be deselected, it may also be minimized using the splitter bars. If the Mouse pointer is in the **Search Spectrum/Spec List Spectrum** Window or the **Hit List Spectrum** window, the RMB (Right Mouse Button) menu option **Change Splitter Orientation** can be used to change the orientation (back and forth between vertical and horizontal) of the divider bar between the plot of and text information for the spectrum. In all cases the RMB menu can be used to access the properties dialog box for each window within the window.

The window on the top left of the **Lib. Search** tab is used for the **Spec List**, and the window on the lower left of this Window is used for the **Hit List**. Note that both the top window and the bar graph of match factors can be minimized using the splitter bars to maximize the hit list space. The window on the left of the **Other Search** and **Names** tabs is used for the **Hit List**, and the window in this position on the **Librarian** and **MSMS** tabs is a list of spectra that can be displayed. It is important to know that the **Spec List** in the **Lib. Search** tab and the **Spec List** in the **Librarian** tab are the same list. Spectra may be copied to the **Spec List**, edited, and deleted from it. The **Hit Lists** are read-only. All of these lists can be displayed as text or as graphic structures by selecting the **Names** or **Structures** tab at the bottom of the window, except for the **MSMS** tab which does not have this alternate display option. Note, only name, match factor, reverse match factor and probability of correct hit are displayed in the graphic mode.

Each **Hit List** (except the one in the **Names** tab) and the **Spec List** can be sorted alphabetically by clicking on the **Name Bar** at the top of the window. The **Hit Lists** in the **Lib. Search** tab and the **Other Search** tab have optionally displayed columns showing the number of synonyms, the number of other databases containing the compound and the retention index. These **Hit Lists** can also be sorted by these columns by pointing to the column header and clicking the left Mouse button.

To copy lines selected in a **Hit List** or **Spec. List** into Windows Clipboard in tab-separated text format, press **Ctrl+K** keyboard keys.

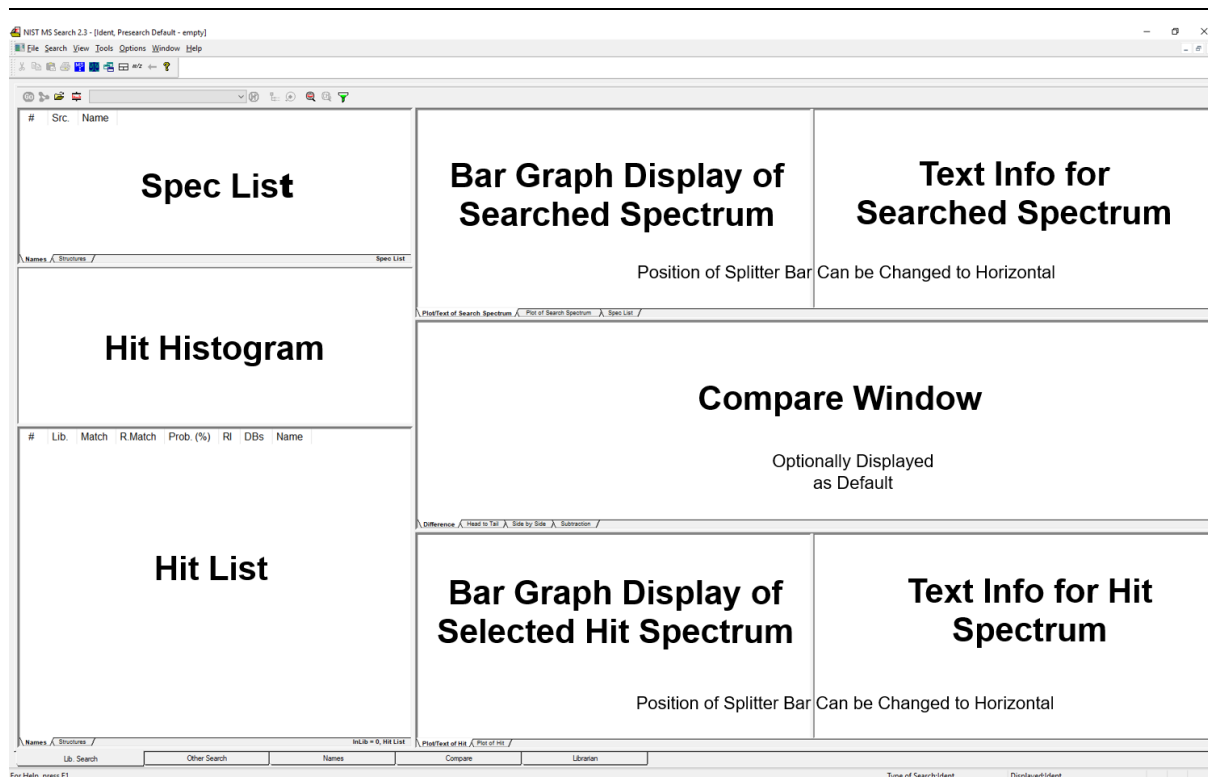


Figure 1. Lib. Search tab.

In some cases, the **Hit Lists** will also have a *plus box* in front of the Hit number. This symbol means that there are replicate spectra for that Hit. Placing the Mouse pointer on the *plus box* and clicking the left Mouse button will cause a list of the replicates to appear below the Hit. Some of the names of the replicates will be preceded by **R** or an **M**, and some by **m** or **r**. If an **R** or **M** precedes the name, the spectrum is of the same compound as the Hit and is in the **NIST replib Library** or the **NIST mainlib Library**. If the name is preceded by an **m** or **r** it means that the replicate is a derivative or isomer of the Hit, with **m** being a spectrum in the **NIST mainlib** and **r** being a spectrum in the **NIST replib**.

The display of spectra of derivatives and isomers as replicates of a compound is controlled through the **Replicates Display Options** dialog box selected from **Options/Replicates** using the **Main Menu** (Figure 2). Details are found by clicking on the **Help** button.

Selecting **View/Derivative Precursor** menu opens **Show Derivative Precursor** for dialog box (Figure 3). Selecting derivative type(s) affects Lib. Search hit list: structures of derivative precursors are displayed instead of derivatives. Only mainlib and replib hits are affected. Deselecting displayed derivative type(s) immediately changes structure display back to derivative. See MS Search help topic "Show Derivative Precursor" for more details.

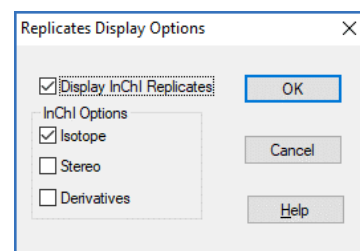
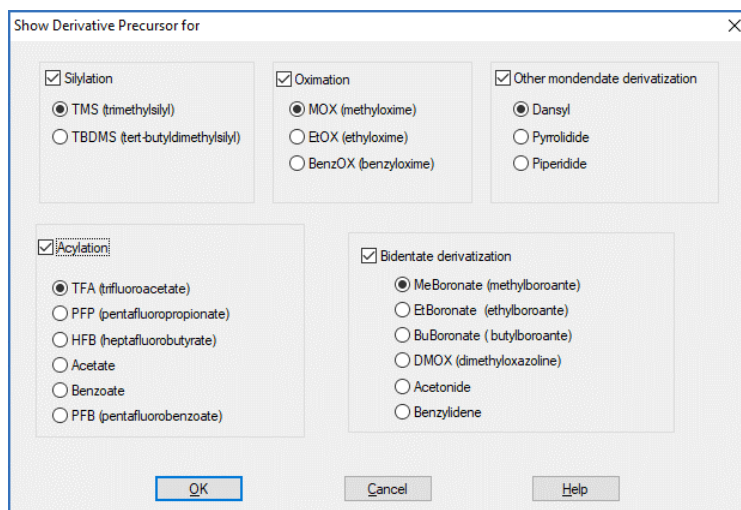


Figure 2

The **Spectral Plot** presentation window (upper right of the Window in the default display) of all tab views except the **Compare** tab, have settable attributes that are found in the **Properties** dialog box described below. By default, the **Plot** window displays the primary name of the spectrum in the lower left-hand corner of the window. In addition, the library from which the spectrum was drawn precedes the name in parenthesis. Optionally, the compound's Molecular Weight (nominal mass), CAS registry number, and Formula (elemental composition) can be displayed. This legend display can be forced to wrap so that all characters will be visible (Figure 4). Also, optional, a structure (if associated with the spectrum) can be displayed on the spectrum. Labeling of the peaks is either by default or the **Plot** tab of the **Properties** dialog box has a "Label peaks greater than (%)" option followed by a *text entry box* allowing for the input

of a number. If the spectrum is to be displayed without peak labeling, enter 100 in this box or deselect **m/z Label** and **Text Label** options. It is also possible to use this **Properties** dialog box to change the colors of the mass spectral peaks, scale, labels, bonds and rings in structures, atomic symbols in structures, and the spectrum's background. Selecting the **Set default layout** button on the **Tool Bar** will restore colors to the original values.



**Figure 3. Select anyone of the five Derivative type check boxes; then select the subcategory radio button.**

The **Text Info** window, by default, is to the right of the **Spectral Plot** presentation in the **Lib. Search** tab. This orientation can be changed so that the **Text Info** window is just below the **Plot** presentation by using the RMB menu selection **Change Splitter Orientation**. In the other tab views where both the plot of the spectrum and the text information are displayed, these two windows are oriented by default with the **Plot** window above the **Text Info** window. Again, this orientation can be changed using the appropriate selection on the RMB menu when the Mouse pointer is on either of the two windows. The **Hit Text Info** tab (**Text Info** tab in some tab views) of the **Properties** dialog box has a number of selectable display options in the form of *check boxes* (**Compound Information**, **m/z intensity list**, **Ten Largest Peaks**, **Synonyms**, and **Retention Indices**). The **Compound Information** is: the primary name associated with the spectrum in its source library; the Formula (elemental composition); MW (nominal mass); Exact Mass (if there is an elemental composition associated with the spectrum); CAS registry number; NIST Number (sequential number assigned to spectra in the NIST EI Archive); ID number (position in the library containing the spectrum); name of the library containing the spectrum; names of Other databases in which the compound can be found, if any (displayed only with **mainlib and replib** spectra); Contributor (NIST Library)/Comments (user libraries); and Related CAS registry numbers, if any (NIST DB only). The other *check boxes* of these first four selections are self-explanatory.

If **Retention Indices** is checked and there are records containing literature GC methods and Retention Indices, these records will be displayed following the display of synonyms (if selected). As pointed out earlier, by default, only the first two records are displayed. If there are more than two records, <...> will appear just below the last line of the bottom record. The **Properties** dialog box allows the setting for the display of a larger selected number of records or all of the records. The record is searchable using the **Find** selection on the RMB menu when displayed in a **Text Info** window. All text is searched so column types, active phases, authors or text in the article titles can be searched. Only the records displayed are searched (**Figure 5**).

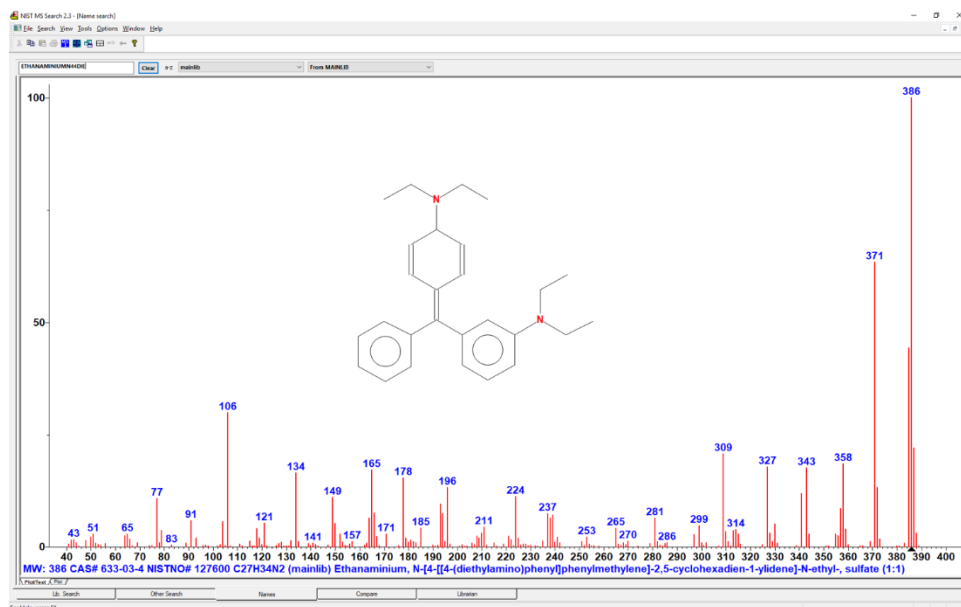


Figure 4. The Spectral Plot display of a spectrum in the Name tab with the nominal mass, CAS registry number, elemental composition name of the library containing the spectrum, and the name associated with the spectrum. The display is wrapped, the structure is shown, and the peaks are not labeled.

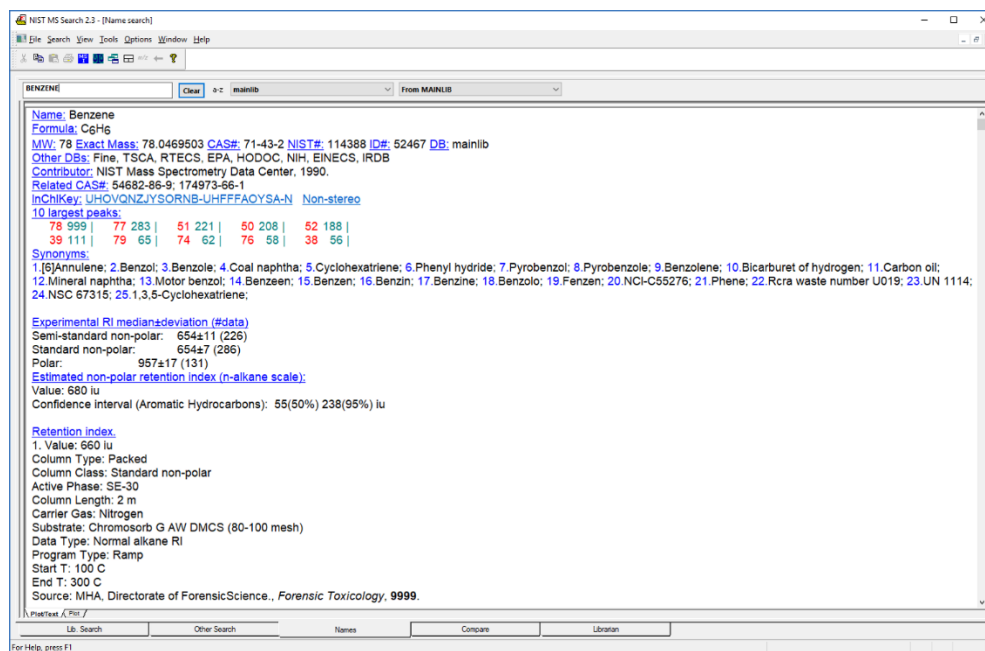
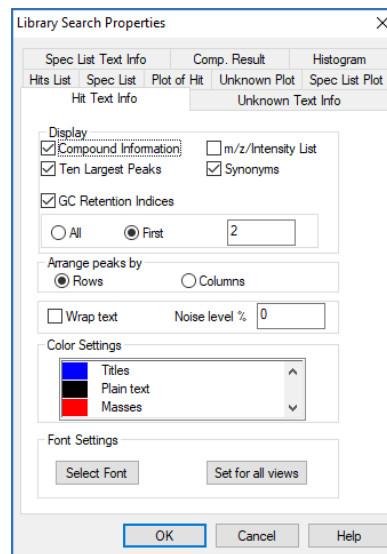


Figure 5. The Text Info display with Compound Information, Synonyms, and Retention Indices as the only selected display options.

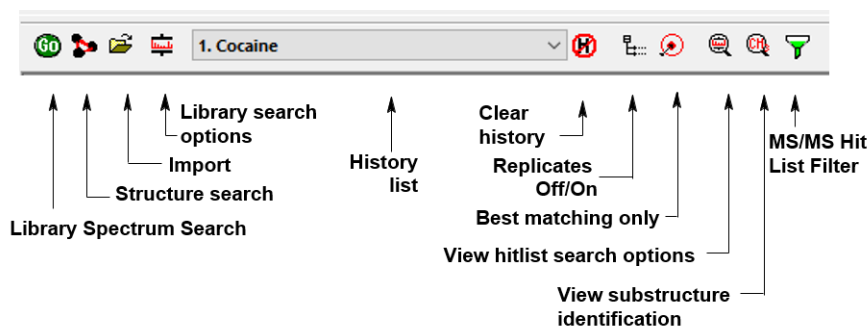
It should be noted that the GC Methods/RI Library is also available separate from NIST 20 with its own Search program, which allows for more extensive use of this Library.

The RMB menu selection **Properties** will result in the display of a multiple-tab dialog box (**Figure 6**). The tabs at the top are self-explanatory. This allows customization of the appearance of the display and control of what is displayed such as a structure on a spectrum, the 10 most intense peaks in the spectrum in the text display, etc. Colors of structures, mass spectral peaks and axes, and fonts used for alphanumeric characters can also be set. Each window can have its own customized set of attributes. This customized display can be saved using the **File/Save Configuration** command from the **Main Menu Bar**. Different saved configurations can be recalled by using the **File/Restore Configuration** command from the **Main Menu Bar**. Separate customizations can be established for each of the six separate tabs. This includes selected libraries to be searched.



**Figure 6. Properties dialog box displayed by selecting Properties on the RMB menu displayed when the Mouse pointer is on one of the windows of the Lib. Search tab.**

At the top left of the **Lib. Search** tab view (**Figure 1**) is the **Button Bar**. The meaning of each of these buttons is shown below:



**Figure 7. Lib. Search's Button Bar.**

The **Go** button will launch a search of the highlighted spectrum in the **Spec List**. The **Structure Search** button will perform a structure similarity search of the highlighted structure in the **Spec List**. The **Import** button (with the file folder icon) is used to import spectra from a text file. The **Library Search Options** button is used to display the dialog box used in selecting libraries to be searched, search types, and other factors associated with the search of spectra and structures.

The drop-down list box contains the **Search History**. A previous search's display can be restored by highlighting an entry in this list. **NOTE:** the **Program** can be set to automatically clear the **Hit List** and/or the **Spec List** by selecting the **Clear History on Exit** check box in the **Hit List** tab and/or **Plot/Text** of the **Spec List** tab in the **Library Search Properties** dialog box displayed by selecting **Properties** from the RMB menu.

The **Clear History** button is used to clear the search history. The **Replicates On/Off** button will toggle the listing of replicate spectra on and off when the **mainlib Library** of the **NIST/EPA/NIH Library** is searched.

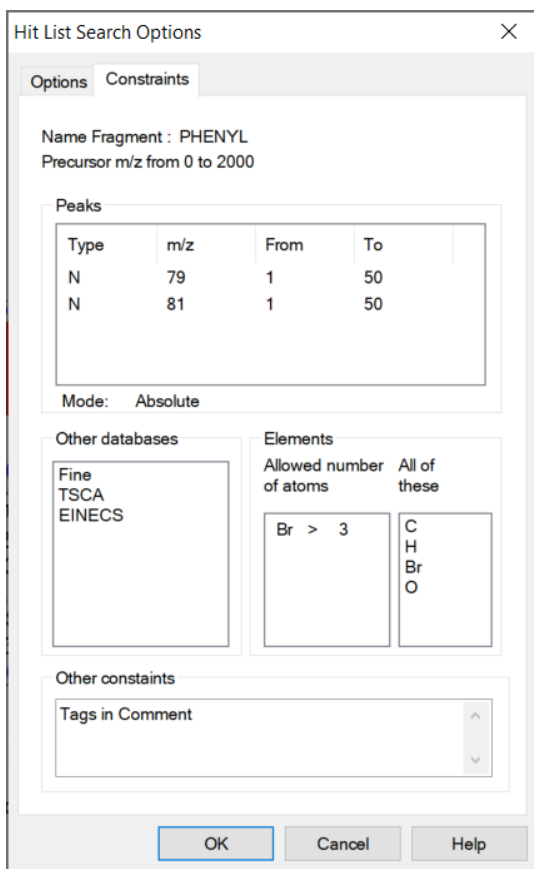
The **Best matching only** button will force the hit list to display only the hit with the highest match factor for any given CASrn. This is the. This feature is especially useful when more than one library is being



searched, and some of the spectra in these libraries have the same CASrn's associated with them. In the case of an MS/MS Search, the best matching spectrum is selected out of Hits that have the same CASrn or names and charge.

The **View hit list search options** button will result in the display of a dialog box with two tabs if constraints have been used in the search. The **Search** or **Options** tab (depending on whether the **Hit List Search Options** are for a **Lib. Search** or an **Other Search** carried out) lists the libraries searched (and type of Search and Search Options for **Library spectrum Searches** only). The **Constraints** tab (if present) lists all different constraints along with values used in the search (**Figure 9** and **Figure 8**).

It should be noted that even though a **Library spectrum Search** is carried out against a list of libraries (**Figure 26**) that does not include **mainlib** and **replib**, these two libraries will be listed as having been included (**Figure 8**) as long as the **Display InChI Replicates** *check box* is selected in the **Replicates Display Options** dialog box (see Figure 2). This will result in the display of the symbol indicating replicates were found when the user-only library was searched. This means that **InChI** search options are available for those spectra in libraries like the Wiley Registry of Mass Spectra and various other libraries provided by Wiley and Others in the NIST format, which have CASrn present in **mainlib**.



Hit List Search Options

Options Constraints

Name Fragment : PHENYL  
Precursor m/z from 0 to 2000

Peaks

Type	m/z	From	To
N	79	1	50
N	81	1	50

Mode: Absolute

Other databases

Fine  
TSCA  
EINECS

Elements

Allowed number of atoms

Br > 3

All of these

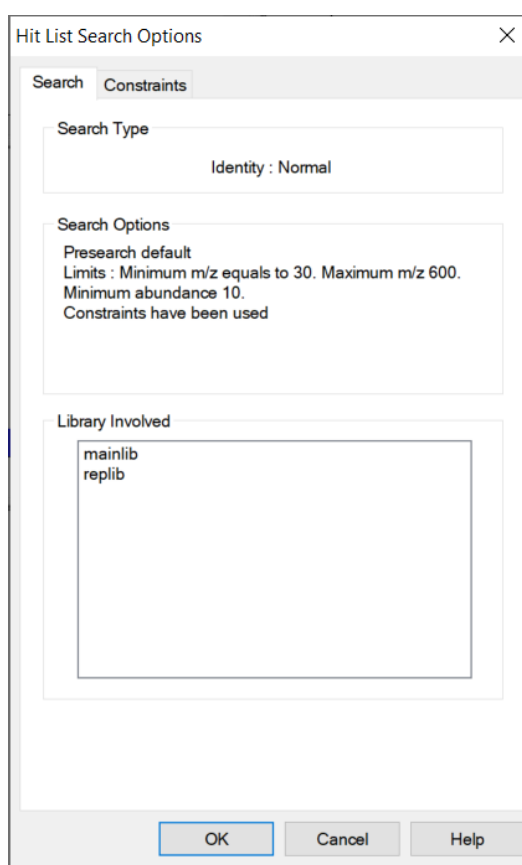
C  
H  
Br  
O

Other constraints

Tags in Comment

OK Cancel Help

**Figure 9.** List of Constraints used for a specific search along with their values.



Hit List Search Options

Search Constraints

Search Type

Identity : Normal

Search Options

Presearch default  
Limits : Minimum m/z equals to 30. Maximum m/z 600.  
Minimum abundance 10.  
Constraints have been used

Library Involved

mainlib  
replib

OK Cancel Help

**Figure 8.** List of libraries used in a Library spectrum Search along with the Search Options

### Two other Important Points about the use of Constraints:

When a **Constrained Search** is carried out, the **Search** is performed as if there was no constraint. Only the spectra that were found in the **Search** that fall within the constraint criteria appear in the **Hit List**.

The number of different constraints that are available exceeds the display area; a slider bar on the right side of the **Constraints** Window in the **Search** dialog box allows access to all constraints. The



---

**Constraint** Window will only allow for the display of six constraint options at a time. Once values have been entered for any of these options, those values will remain until individually changed or deleted. Once an available Constraint has been selected (check mark in *check box* entered by putting the Mouse pointer on the *check box* and clicking the left Mouse button), the values previously entered for that constraint will be used when a **Constrained Search** is performed. Because if a Constraint is selected and is not visible in the Window, a bad result may occur, the program displays the number of constraints selected. A **Clear All** button has been added to the **Constraints** tab of the **Search** dialog box. The **Clear All** button removes the checks in the various constraints, it does not change the values in each of the constraints. It is recommended to click on this button before starting a new **Constrained Search** and reselecting any desired Constraints. Whenever any Constraint is selected, the **Use Constraints** *check box* is automatically selected. Clicking on any individual Constraint's *check box* in the **Constraints** Window causes appearance of the **Constraints** tab of the **Search** dialog box to change to allow for the entry (or edit) of values for that Constraint at the same time the Constraint is selected or deselected. The appearance can also be changed to that for a specific constraint by clicking on the constraint description away from the *check box*.

The last button on the right, the **View substructure information** button, displays the **Substructure Information** dialog box. The **Substructure Information** dialog box contains a list of the probability of the presence and absence of substructures that may be associated with a compound that produced the searched mass spectrum along with additional predictions about this compound's molecular weight and estimate of the numbers of atoms of chlorine and/or bromine that may be contained. This is especially useful with a **EI Hybrid**, **EI Simple**, **EI Neutral Loss**, **MS/MS in EI**, or **MS/MS Hybrid Similarity**, or **MS/MS In EI Similarity Search**. An example of the use of **Substructure Information** resulting from an **MS/MS in EI Similarity Search** of the **NIST/ERP/NIH EI Library** is provided later in this manual.

#### **Title Bar, Tool Bar, and Status Bar**

At the top of the display of each of the **Program's** tab is a **Title Bar**. The contents of a title bar of the active tab are also displayed in the **Title Bar** of the **Program** Window. When the **Lib. Search** or **Other Search** tab is displayed, this **Title Bar** will contain information about the currently displayed search. In the case of a **Library spectrum Search**, this will include the search type used, whether or not constraints were used, and if any type of limits were applied to the search. In the case of the **Other Search** tab, the **Title Bar** will contain the type of search that was done (i.e., **Formula**, **Molecular weight**, **Any peaks**, etc.) and whether or not constraints were used. When the **Compare** tab is displayed, the **Title Bar** contains the number of spectra available for comparison regardless of the number displayed.

At the bottom of the **Program's** display, below the **Program's** tabs, is the optionally displayed **Status Bar**. The **Status Bar's** display can be turned off by selecting or deselecting **Status Bar** from the **View** menu. When a **Library spectrum Search** is being conducted, the left side of the **Status Bar** will have displayed "Comparing XXX Library spectra with submitted spectrum" where XXX is the number of spectra that were found during the Presearch (see **Search Algorithms** in Appendix 5). If **Off** has been selected in the Presearch area of the **Search** tab of the **Library Search Options** dialog box, XXX will represent the total number of spectra in the libraries being searched.

There are two fields on the right side of the **Status Bar** when the **Lib Search** tab view is active. The left-most of these five fields will display the search type that is selected in the **Library Search Options** dialog box following the **Type of Search:** label ((EI) **Hybrid**, (EI) **Simple**, (EI) **Neutral Loss**, **MS/MS in EI**, or **MS/MS Hybrid**) if the **Similarity Search** is selected; and (EI) **Ident**, (EI) **Quick**, **MS/MS**, (In-source) **HiRes**, depending which **Identity Search** has been selected).

The next information shown following the **Displayed:** label is the type of search responsible for the current display. These are the same as the **Types of Search:** types. The will contained **Rev** if the

**Reverse Search** *check box* with selected in the **Search** tab of the **Library Search Options** dialog box. (the Match Factor obtained by ignoring all peaks that are in the sample spectrum but not the library spectrum). The area where the **Rev** label appears will contain the number used as a precursor Precursor MW value in a **Neutral Loss** or **Hybrid Similarity Search**. If an **Identity Search** has been carried out, another label will appear to the right (**Penalize**) if the **Penalize rare compounds** *check box* was selected in the **Spectrum Search Options** section of the **Search** tab of the **Library Search Options** dialog box.

The **Tool Bar's** display (Figure 10) is optional by selecting it from the **View** menu. The functionality of its button **Switch to Caller** and **Print Report** is also available on the **File** menu; **MS Interpreter**, and **AMDIS** also may be launched from the **Tools** menu.

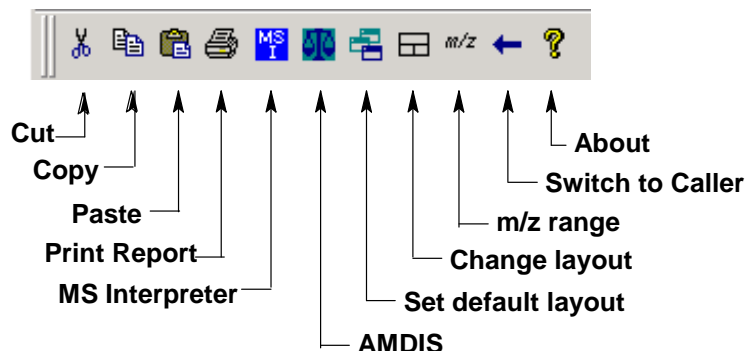




Figure 10. NIST MS Search Tool Bar.

A Tool Tip is associated with each of the buttons with an explanation of the button's function. The  button will return all displays on the currently viewed tab to the default values. The  button toggles through three separate views for the currently displayed Program's tab.

The **m/z range** button displays a dialog box (Figure 11A) to fix the starting and ending *m/z* values for bar-graph spectra displays, and to choose whether the peaks should be labeled with neutral losses or with their *m/z* value.

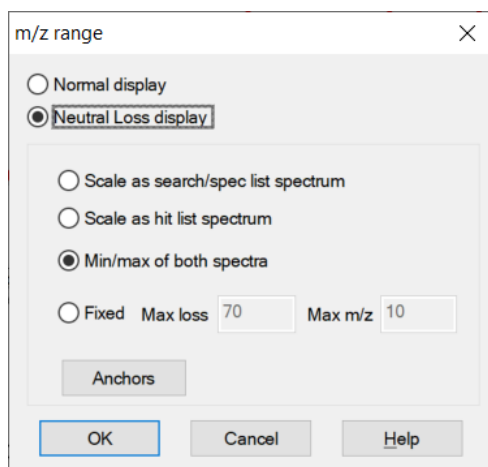


Figure 11A : *m/z* range dialog box 1

An explanation of each of the items in this dialog box can be found by selecting the dialog box's **Help** button. The **Anchors** dialog box is used when viewing spectra with a neutral loss label. Selecting **Anchors** will result in the display of the **Anchors** dialog box, which is explained by selecting the **Help** button in that dialog box.

The two radio buttons at the top of the ***m/z* range** dialog box have to do with labeling spectra normally or with natural loss values. **Normal display** is the most common.

## Neutral Loss Display

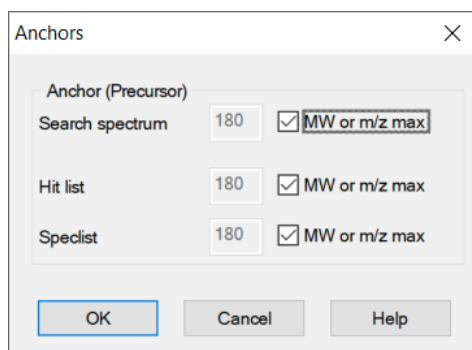


Figure 11B. *m/z* range (for Spectra Display) dialog box.

One of the important tools in deducing a structure from a mass spectrum is to look at the mass spectral peaks as losses from the molecular ion rather than as individual peaks. In this way of viewing the mass spectrum, the peak at *m/z* 269 in methyl stearate becomes the  $[M - 29]^+$  peak (loss of an ethyl radical) and the peak at *m/z* 267 becomes the  $[M - 31]^+$  peak (loss of a methoxy radical). See Figure 12.

The program can display a spectrum with neutral loss labeling with the library spectra and search spectrum be labeled as neutral losses with the first label on the right-most peak being 0 (Figure 13). To activate this display, select **Neutral Loss Display** from the RMB menu. Selecting **Neutral Loss Anchors** is explained in the Help files (topic *m/z* range). NOTE: If you do not know the molecular weight of the search spectrum, the program will attempt to estimate it. This value may be overridden by using the librarian to edit the spectrum and specifying a molecular weight and then using the edited spectrum to run the library search.

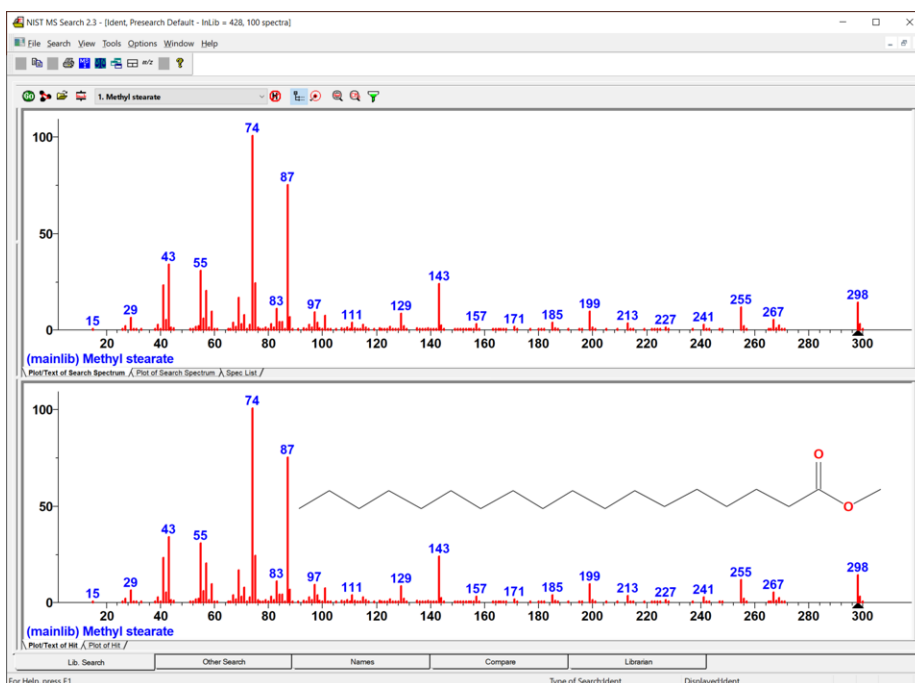


Figure 12. Nominal spectrum labeling where the top spectrum is of an unknown compound and bottom spectrum is of the first Hit from a search of the NIST/EPA/NIH Mass Spectral Library (NIST 20).

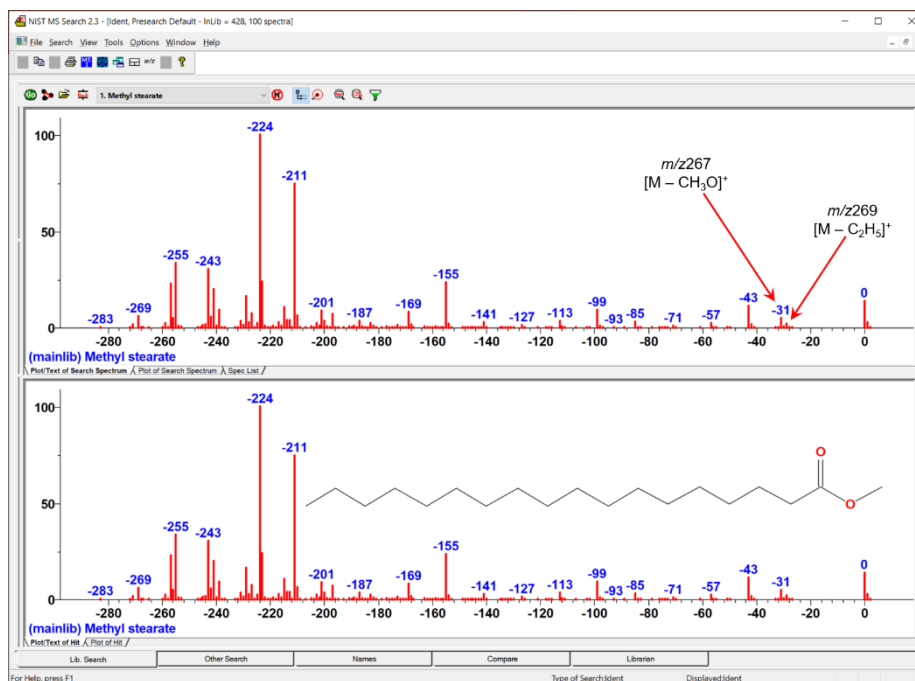


Figure 13. Neutral Loss spectrum labeling where the top spectrum is of an unknown compound and bottom spectrum is of the first Hit from a search of the NIST/EPA/NIH Mass Spectral Library (NIST 20).

## The Other Search Tab

The second tab display, **Other Search** (Figure 14), is used to perform a **MW (Nominal Mass)**, **Exact Mass**, **MS/MS Precursor  $m/z$** , **Formula** (elemental composition), **Any Peaks**, **CAS registry number**, **NIST Number**, or **ID Number Search**. The **NIST Number Search** can only be carried out on the **NIST Main Library (mainlib)**, the **Replicates Library (replib)**, the **Tandem Libraries** and the **Retention Index Library (nist\_ri)**. **NIST Numbers** establish correspondence between spectra in different releases of the **NIST/EPA/NIH MS Library** and can be carried out on one or all of these libraries at a time. The **ID Number Search** can only be carried out on one library at a time and can be used with any of the NIST libraries or other libraries such as User Libraries. The **MS/MS Precursor  $m/z$  Search** works with the NIST Tandem libraries and user libraries containing Tandem spectra. The other options in this group can search multiple libraries (maximum of 127), and different sets of libraries can be specified for each different type of search.

As shown in Figure 15, a feature introduced in v.2.2 of the **Program** was the display of columns showing not only the number of synonyms, the number of other databases (non-mass spectral databases) in which the Hit can be found but also the retention index (RI) if available. Not only are these columns available for the **Hit List** of searches performed in the **Other** tab, but also in the **Hit List** displayed in the **Lib. Search** tab. The display of these columns is controlled by the **Hit List** tab of the **Properties** dialog box displayed by selecting the **Properties** option from the RMB menu displayed when the Mouse pointer is in a **Hit List** window.

The **Any Peaks Search** is a powerful tool which is explained in more depth later in this manual and in the Program's Help File. This has been expanded to allow for peaks with accurate  $m/z$  assignments to be searched against the NIST and user libraries.

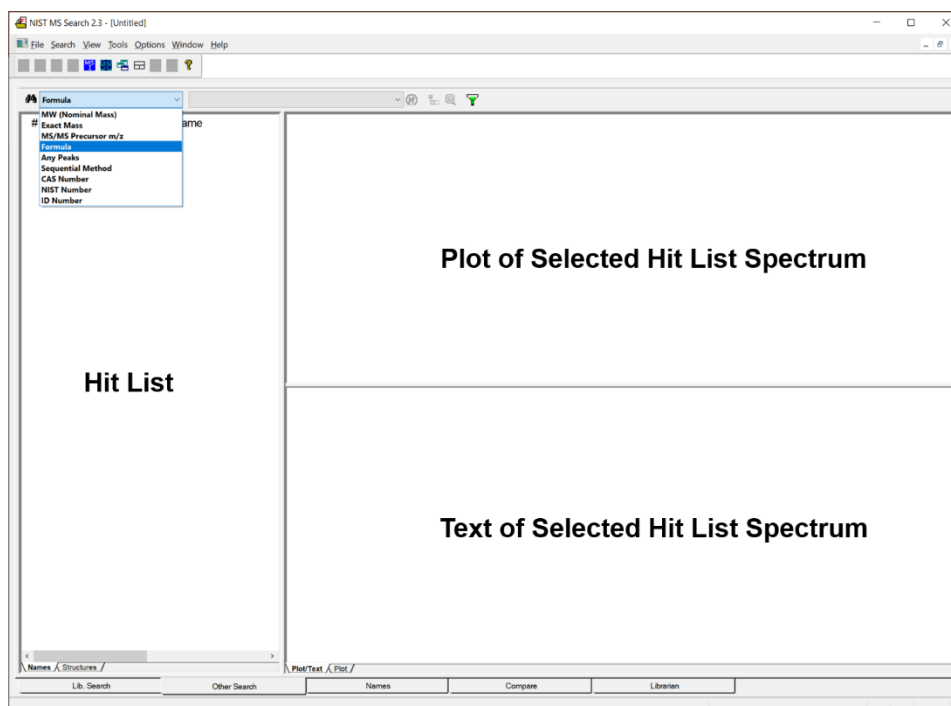


Figure 14. Other Search tab.

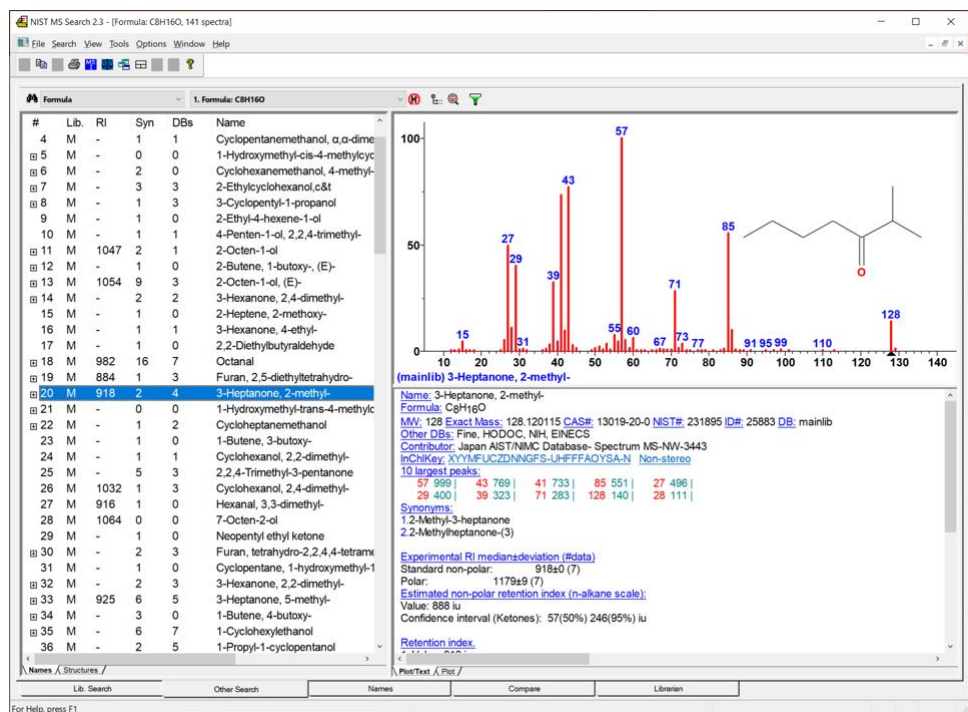


Figure 15. Results of a Formula Search showing the number of “synonyms” for the Hits, the number of other “databases” in which the Hits can be found and the RI value (if present) columns.

Another very import search carried out in this tab view is the **Sequential Method Search**. This is a wild-card search which allows the user customer the mass spectral and meta data included in the search. With the **Sequential Method Search**, the search can even be limited to only spectra that come from a specific contributor.

## The Names Tab

The third tab, **Names**, is used for the NIST signature **Incremental Name Search (Names)** (Figure 16). The **Name Search** can only be performed against one mass spectral library at a time. The library to be searched is selected by use of the **Library to search** drop-down list box next to the **a-z** button. If the **a-z** button is selected, the only input in the text entry field on far left just below the button bar that will be taken is that of letters **a-z**. If the **a-z** button has not been selected, numbers as well as letters can be input. Names containing Greek alphabet characters can be pasted into the text entry field. Typed dashes and other punctuation characters will not be used. All letters are entered as upper case regardless of the input.

NOTE: The drop-down list box just to the right of the **a-z** button is used to select the library to be searched. This search method allows for the **browsing** of user libraries.

The second list box only appears when the library selected is one of the NIST mainlib is searched. It is used to display replicate spectra for the **NIST 20 mainlib Library**. If there are no replicate spectra for the highlighted selection in the **Hit List**, this list box will not appear on the display. When Replicates are available they will appear in the list box as either **Replicate** followed by a sequence number or **\*Replicate**. **Replicate** is a spectrum of the same compound that is in the **replib Library**. **\*Replicate** identifies a spectrum that is of a heavy isotope variant of the compound, a stereoisomer of the compound, or a compound, which has the same derivative precursor.

The **Name Search** used with the **NIST/EPA/NIH Mass Spectral Library** uses not only the main name but also any synonyms. As an example, the **mainlib** compound name for acetylsalicylic acid is the common name *Aspirin*. This compound has 157 synonyms and its spectrum can be displayed by entering any of these names including trade names like *Bayer* (Figure 16).

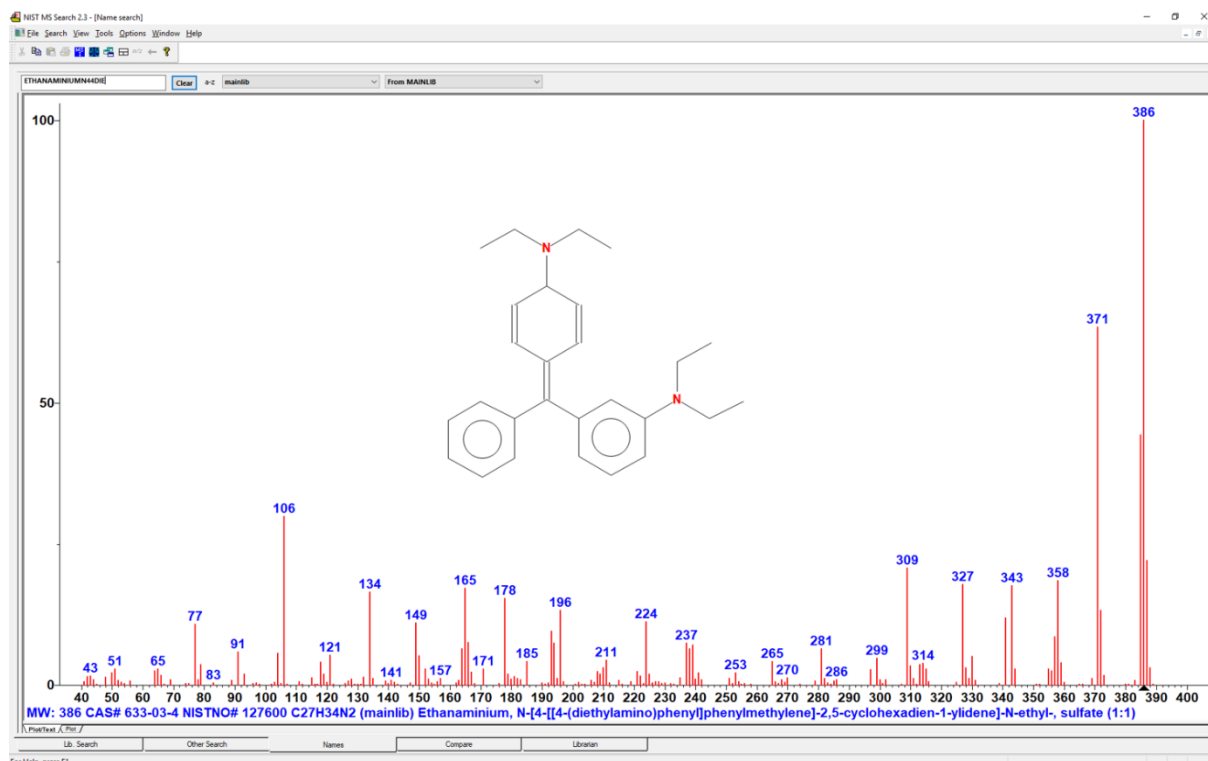


Figure 16. Incremental Name Search tab.

Note, the use of synonyms is also supported in user libraries. Thus, if a spectrum is given a synonym "Lab Book 4-PG-22", then the data can be retrieved by that name using the incremental search.

## The Compare Tab

The fourth tab, **Compare** (Figure 17), allows the display of a sample spectrum and multiple spectra from the **Hit List**. Other spectra can also be put into the **Compare** display. A selected spectrum in the lower part of the **Compare** tab can be displayed as various types of comparisons with the spectrum in the upper part of the display by selection of one of the tabs in the **Compare** window (**Difference**, **Head-to-Tail**, **Side-by-Side**, **Subtraction**).

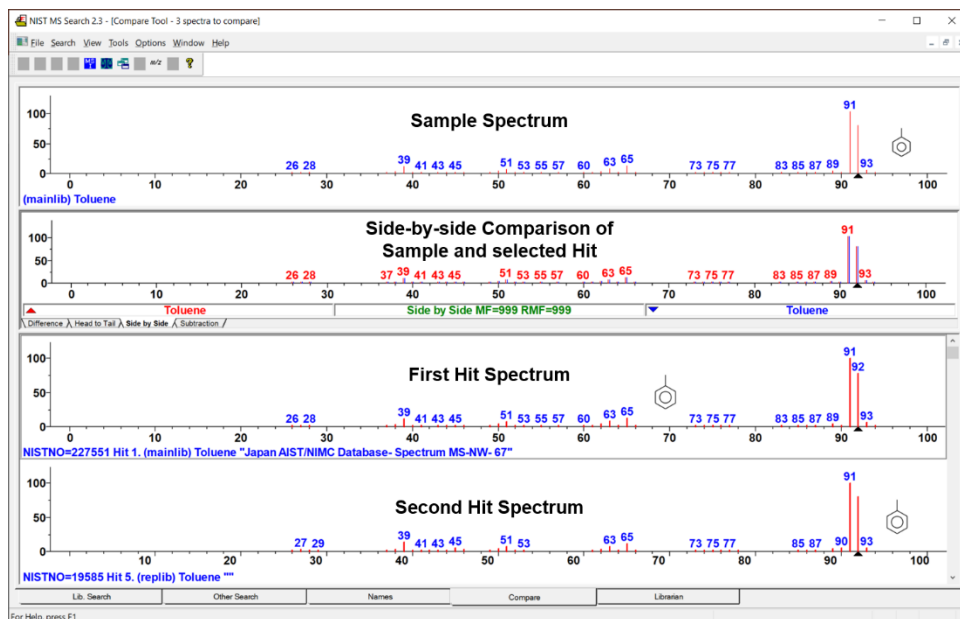


Figure 17. Compare tab.

## The Librarian Tab

The fifth tab, **Librarian** (not shown), is used to deal with spectra that are in user libraries or that are to be added to user libraries or have been imported or searched. The layout of this tab is the same as the tab used for the **Incremental Name Search** with the exception that the far-left window contains the **Spec List** rather than a **Hit List**. This is the same **Spec List** as displayed in the upper left window of the **Lib. Search** tab. The **Librarian** tab allows for the editing or creating of spectra. This editing not only allows  $m/z$  intensity pairs to be edited, deleted, or added; but also allows CAS registry numbers, elemental compositions, structures, names, synonyms, and structures to be edited, added, or deleted. In addition, detailed text information about the spectrum can be added using a predefined set of titles.

When a splitter bar divides a window, the position can be changed by selecting **Change Splitter Orientation** on the RMB menu, displayed when the pointer is in the window.

**Zoom.** Placing the Mouse pointer on a mass spectrum plot or spectra comparing bar-graph plot, holding the left Mouse button down and dragging the Mouse pointer along a diagonal will result in drawing a zoom rectangle. After releasing the Mouse button, a zoomed spectrum will be displayed; the cursor will change to a magnifying glass. To return to the normal display, select **Zoom Out** from the RMB menu or double-click.

Additional information on the **Librarian** and **MSMS** tabs is provided later in this manual under **User Libraries** and **MS/MS Searching**.



---

## Spectral Analysis Utilities for NIST 20

### ***Automated Mass Spectrometry Deconvolution and Identification System (AMDIS)***

Included as a separate utility, **AMDIS** attempts to reconstruct original mass spectra for individual components in arbitrarily complex GC/MS and LC/MS chromatograms; and if a target library is provided, **AMDIS** can directly identify target compounds. **AMDIS** is especially useful when a single reconstructed total ion current (RTIC) chromatographic peak represents multiple components. Regardless of each component's concentration, pure mass spectra are deconvoluted for analyses. **AMDIS** was developed by NIST under a contract from the Defense Threat Reduction Agency (DTRA), Department of Defense, for verifying of compliance with a major international treaty (Chemical Weapons Convention) as ratified by the United States Senate in 1997. To meet the rigorous requirements for this purpose, **AMDIS** was tested against more than 30,000 GC/MS data files accumulated by the EPA Contract Laboratory Program without a single false-positive for the target set of known chemical warfare agents. Although this level of reliability may not be required for all laboratories, this shows the degree to which the algorithms have been tested. **AMDIS** can also be used to identify individual non-target components in complex samples. It has been widely used in the flavors and fragrance industry, forensics, the study of natural products, and for the identifications of metabolites.

**AMDIS** has been designed to reconstruct "pure component" spectra from complex RTIC chromatograms even when components are present at trace levels. For this purpose, observed chromatographic behavior is used along with a range of noise-reduction methods. **AMDIS** is distributed with specialized libraries (environmental, flavor and fragrance, and drugs and toxins) that were derived from the NIST 02 MS Library. **AMDIS** has a range of other features including the ability to search the entire **NIST 20 MS Library** with any of the spectra extracted from the original data file. It can also employ retention-index windows when identifying target compounds and can make use of internal and external standards maintained in separate libraries. A history list of selected performance standards is also maintained.

The instruments that are supported in their native file format are: †

- Agilent ChemStation (GC-MS & LC-MS) and MS Engines
- BrukerDALTONICS GC-MS (\*.MSF)
- Finnigan (GCQ, INCOS, ITDS, ITS-40, ITD-700/800, ITMS)
- Inficon (\*.acq)
- JEOL/Shrader (\*.lrp)
- Kratos Mach3 (\*.run)
- Micromass (MassLynx, NT Formats) (RAW)
- mzXmL/mzData
- Perkin-Elmer TurboMass (RAW) and ITD-700 (\*.ms)
- Shimadzu MS Files QP5000 and QP5050 (\*.R##)
- Thermo Fisher Xcalibur (RAW)
- Varian (MS, SMS, XMS)

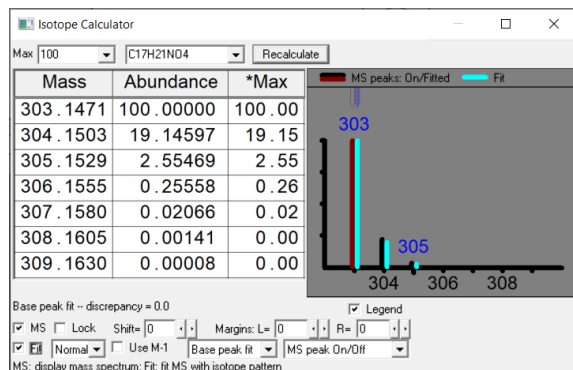
In addition, the NetCDF format, which is supported by many manufacturers, is supported by **AMDIS**. Other formats will be added in future releases. Contact NIST for future information.

† ***The use of specific product and trade names does not indicate an endorsement by the National Institute of Standards and Technology. The specific file formats are supported in order to ensure that the data is transferred to AMDIS with no information loss. The specific formats and names are covered by the respective copyrights of the individual companies.***

## MS Interpreter

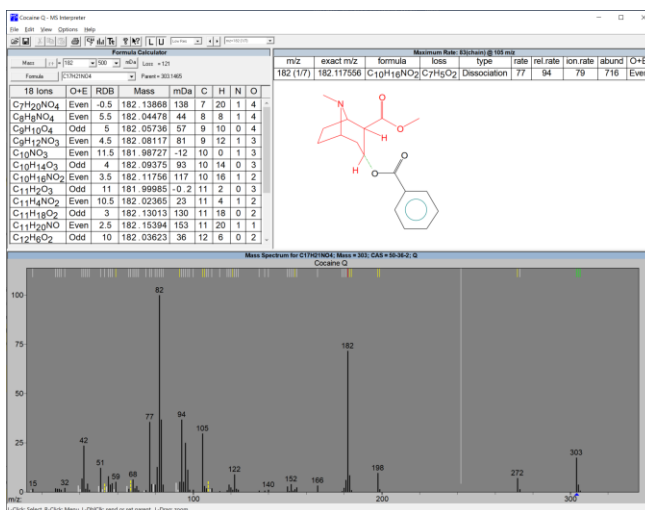
This tool was developed to aid NIST evaluators in their analysis of mass spectra. In one integrated program, it enables a variety of calculations on a mass spectrum, using a proposed chemical structure if available. Spectra and structures are associated in the spectral editing component of the **Librarian** tab of the **MS Search Program** discussed in **USER LIBRARIES** section and **APPENDIX 3** of this manual. **MS Interpreter** is activated for an individual spectrum with the Mouse Pointer on a **Plot** or **Text Information** window or by highlighting the name or structure of a compound in a **Hit List** and selecting **MS Interpreter** on the **Send To** menu of the RMB menu. This action opens **MS Interpreter** with the spectrum and associated structure (Figure 18). **MS Interpreter** can also be opened from the **Tools** menu or a button on the **Tool Bar**. If a spectrum/structure has been opened in **MS Interpreter** and then the program closed and then reopened from either the **Tools** menu or the button on the **Toolbar**, the last contents will be displayed. If **MS Interpreter** is started by either of these two actions just after starting the **Program**, both the **Structure** window and the **Spectrum** window will be empty. Spectra can be imported from MSP files (text files containing spectra in the **NIST MS Search** format), JCAMP-DX, and SDF files. If the folder also contains a \*.STB structure file that has the same name as the MSP file, that structure will be associated with a spectrum in **MS Interpreter**. When a structure/spectrum pair is Saved from the **MS Interpreter's** File Menu, a single SDF file containing the spectrum and the is created.

Peaks in the spectrum originating as a logical fragment of the molecule ( $M^{+*}$  or  $MH^{+}$ ) are marked at the top of the spectrum display; and corresponding fragments, derived from thermodynamic approximation, may be highlighted by clicking on one of these marks. It also allows for keeping track of important neutral losses, both from the precursor or a derived ion and to find all possible formulas for any peak or neutral loss. Isotopic patterns (Figure 19) for any cluster can be separately examined. These isotope clusters can be compared to theoretical predictions subject to a number of user-specified constraints. **MS Interpreter** is a separate program that also has an icon in the **NIST Mass Spectral Database** program folder. It also has a detailed Help screen that can be called from the program by selecting **Help** on its **Menu Bar** or by pressing the F1 key. Information about the control currently located under the mouse



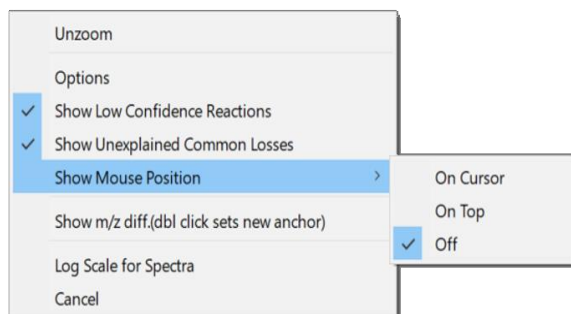
**Figure 18.** View of MS Interpreter's isotope calculator used to compare observed abundances to theoretical values.

pointer and possible user actions is displayed in the **Status Bar** located at the bottom of the program's main Window. Placing the Mouse Pointer in the mass spectrum display and clicking the Right Mouse button will result in a somewhat self-explanatory RMB menu. Before using the program, explore this RMB menu (Figure 21).



**Figure 19.** The main Window of MS Interpreter. The black lines indicate the peaks in the mass spectrum that the MS Interpreter can explain. The white lines indicate the ones it cannot. The color coding is settable by the user to allow specific types of ion formation mechanisms to be shown.

Selected ions on the plot of the mass spectrum can be sent to **Formula Calculator** and the **Isotope Calculator** using selections on this menu. A specific  $m/z$  value can be locked as the position to display calculated losses from. In addition, the **Fragment options** dialog box can be displayed (Figure 20).



**Figure 21. Right Mouse button menu displayed when the Right Mouse button is on the Plot portion of the MSI Program.**

More detailed information can be obtained about the use of this utility by pressing the **F1** key with the Mouse pointer in one of the **Program's** windows.

#### Exact masses of fragment ions

Structure view displays not only nominal  $m/z$  but also exact mass of the predicted fragment ion.

#### Pasting a chemical structure into the Structure view

- Copy a structure from a structure editor (molfile format) into Windows Clipboard;
- In MS Interpreter, click Structure view title bar to activate it;
- Click **Paste** button or select from the menu: **Edit/Paste** or press **Ctrl-V**.

#### Pasting a mass spectrum or a chemical structure into Mass Spectrum view

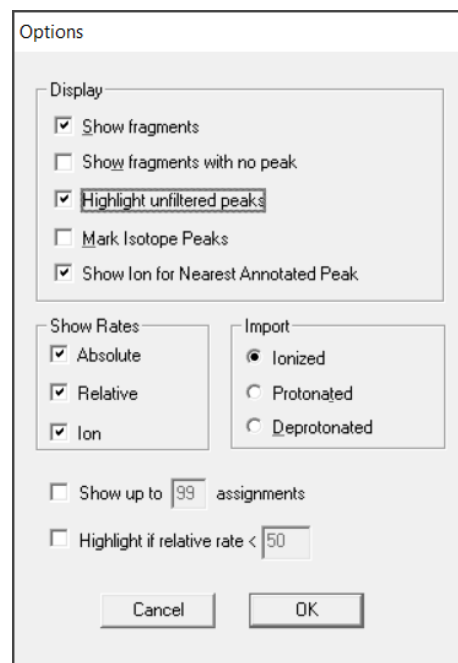
Contents of a file containing a mass spectrum or a structure or both (SDfile) may be pasted into the Mass Spectrum view.

#### Pasting a peak list into the Mass Spectrum view

- Copy a peak list into the Windows Clipboard;
- In MS Interpreter, click **Mass Spectrum** Window to activate it;
- Click **Paste** button, or select from the menu: **Edit/Paste**, or press **Ctrl-V**.

#### Notes

- the peak list must be made out of  $m/z$  and intensity pairs separated by any of the following characters:  
**space , ; : ( ) [ ] } | Tab**
- each line must contain one or more  $m/z$  and intensity pairs;
- if at least one line contains an odd number of values, then the program reads only first two numbers (one peak) from each line; a warning is displayed;
- all peaks must be in order of either ascending or descending  $m/z$ ;
- all  $m/z$  are rounded to the nearest integer by the MS Interpreter;



**Figure 20. Fragment options dialog box.**

- 
- there is no need to normalize intensities to 999.

### Pasting MSP, SDF, MOL, and JDX (JCAMP-DX) files from Windows Explorer

- MSP, SDF, MOL, and JDX files may be copied from the Windows Explorer and pasted into the Structure or Mass Spectrum view of MS Interpreter.
- The files of the same type may be dragged from the Windows Explorer and dropped into the same MS Interpreter views. *In all cases, only the first spectrum and/or structure from the first selected file is/are imported.*
- The name of the imported file appears on the MS Interpreter **Title Bar** and in the MRU (most recently used) list in the **File** menu.

### Copying calculated isotope envelope table from the Isotope Calculator into Windows Clipboard

- Right-click any line in the Isotope Calculator table
- Select **Copy**.

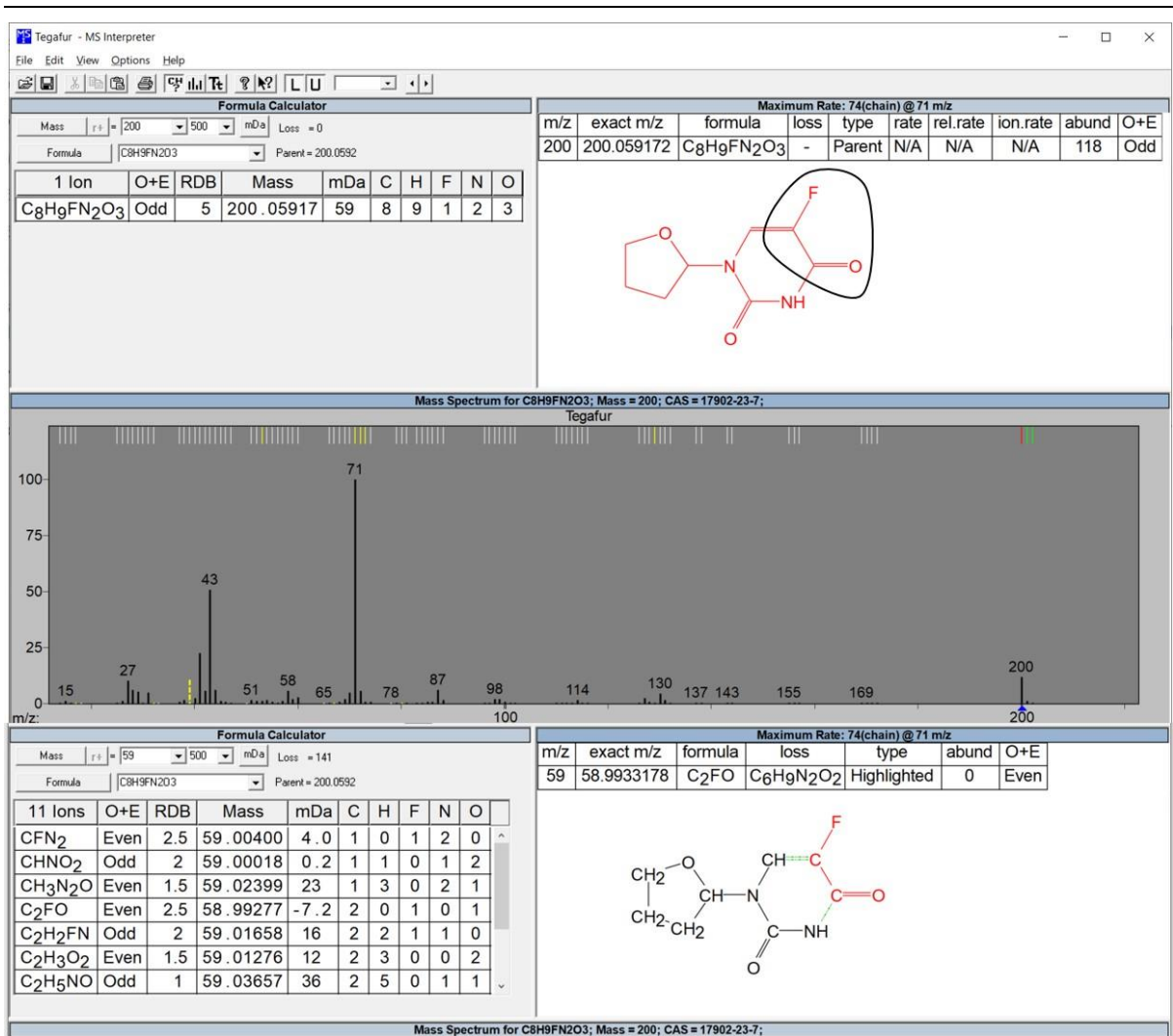
This will copy into the Windows Clipboard a tab-delimited text, which is suitable for pasting into a MS Excel spreadsheet or text editor. The context menu and the text contain words “Default Isotope Table” or the name of the file from which a custom isotope composition was imported (see the next section.)

### Using custom isotope composition table in Isotope Calculator and Formula Calculator

- Open file *MS\_Interpreter\_isotope\_table\_test\_01.itbl* in a plain text editor. This file is installed in the same folder as MS Interpreter, typically, c:\nist20\mssearch
- Read the notes in the file
- Edit the file and save it under another name
- From **MS Interpreter** menu, select **File/Import Isotope Table**
- In the **Open Isotope File** dialog box, select the newly created file and click the **Open** button.
- To restore the original table, import file *MS\_Interpreter\_isotope\_table\_default.itbl*. The original table is restored automatically after both MS Interpreter and MS Search are shut down.
- Every time a \*.itbl file is imported, a \*.itbl.log file is created or appended. This file contains the date, time, a full pathname of the .itbl file, and a list of differences (if any) between the current and the imported isotope table or a reason for the failure to import the new table.

### Selecting a Portion of the Structure for Mass Calculation (lasso function)

You may drag the mouse around the displayed chemical structure to compute the mass of the selected substructure (highlighted in red). Put the Mouse pointer on a starting point in the Structure window. Hold down the left Mouse button as you drag the pointer around the portion of the structure you want to select until the pointer is back at its starting position (lasso the portion of the structure you are interested in). When the left Mouse button is released, information is updated, and the Structure window shows the selected portion of the structure in red. The mass will also be shown on the spectrum. This need not correspond to an actual peak – predicted or observed. Illustrated below.



This Figure shows 'Lassoing' part of a structure (top) in the MS Interpreter display and the change in the display bottom. The Formula Calculator (bottom) is automatically updated.

---

## Example Searches

### *Library Spectrum and Structure Searches*

The **Program** can be used to find spectra that are similar to that of an unknown compound. This can be accomplished by importing the spectrum from a third-party mass spectral data analysis program, from a text file (\*.MSP), or from an MSP file that is on the Windows Clipboard. Many instrument manufacturers (like Agilent, PerkinElmer, Waters, LECO, ThermoFisher and others) have utilities in their data analysis software that creates files in the MSP format, starts the NIST Search Program or brings it into focus, and sends the spectrum (or a group of spectra) to the Program which will automatically import them and run the search, if so configured. Lacking this link in the manufacturers software, some commercial systems can output a data file in the netCDF format. This file can be imported into AMDIS and spectra can be exported for automatic search directly to the MS Search Program. MS Search can also import in the idx, SDF, KCF (Glycans), MGF, DAT, PKL, mzXML, and mzData formats.

A spectrum can be created using the **Spectrum Information** dialog box displayed by selecting the **New** button on the button bar of the **Librarian** tab to enter the  $m/z$  and intensity pairs that make up a mass spectrum. The **Library Spectrum Search** and the **Structure Search** are usually carried out from the **Lib. Search** tab by clicking on the **Library Search** button (**Go**) or **Structure Search** button (next to the **Go** button). Another way of initiating a search is to select a spectrum in any **Hit List** or the **Spec List**, **Plot** window or **Text Info** window in any tab, or in a **Compare** window; display the RMB menu and **click with the Right Mouse button any spectrum** in a List of spectra or Subtraction spectrum in the **Compare** window; and select the desired search (**Library Search** for a spectrum search or **Structure Similarity Search** for a structure search). To search for more than one spectrum at once, make a multiple selection in any List of spectra (**Hit List**, **Spec List**, or **Other Search List** except **Name Search List**), display the RMB menu, and select **Library Search** or **Structure Similarity Search**. Regardless of the tab display, when one of these searches is carried out, the view will change back to **Lib. Search** tab. In the default display of the **Lib. Search** tab, the **Hit List** is on the lower left. The third column (from the left) of the **Hit List** contains the Match Factor. Match Factor 1000 for the **Structure Search** does not mean an exact match; it means that the found structure is a homologue of the submitted structure. Match Factors for the **Spectrum Search** cannot exceed 999.

Spectra and structures can be imported from files or third-party programs through various types of automations, described later in this section. Spectra or structures can be imported from a file by selecting **File** from the **Main Menu Bar** and selecting **Open** from the **File** menu (Figure 22). Selecting a file will result in a display of the **Structure Import** dialog box or the **Spectrum Import** dialog box. The **Structure Import** dialog box is self-explanatory. The **Spectrum Import** dialog box (Figure 23) is more complex. When a file containing more than one spectrum is selected, one or more selected spectra or all of the spectra can be imported by using either the **Import Selected** or the **Import All** button, respectively. Pressing the **Import Options** button will open the **Spectral Import Options** dialog box (Figure 24). This is where information about the  $m/z$  values of peaks is controlled.

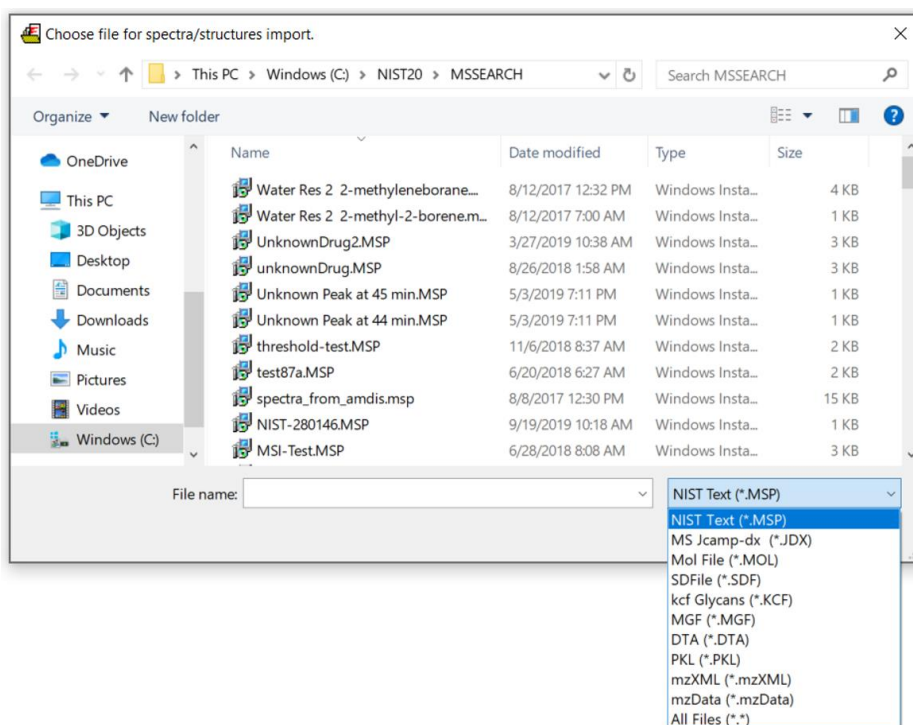


Figure 22. File Selection dialog box with some of the available file types. MSP and JDX files are spectra, and MOL files are structures.

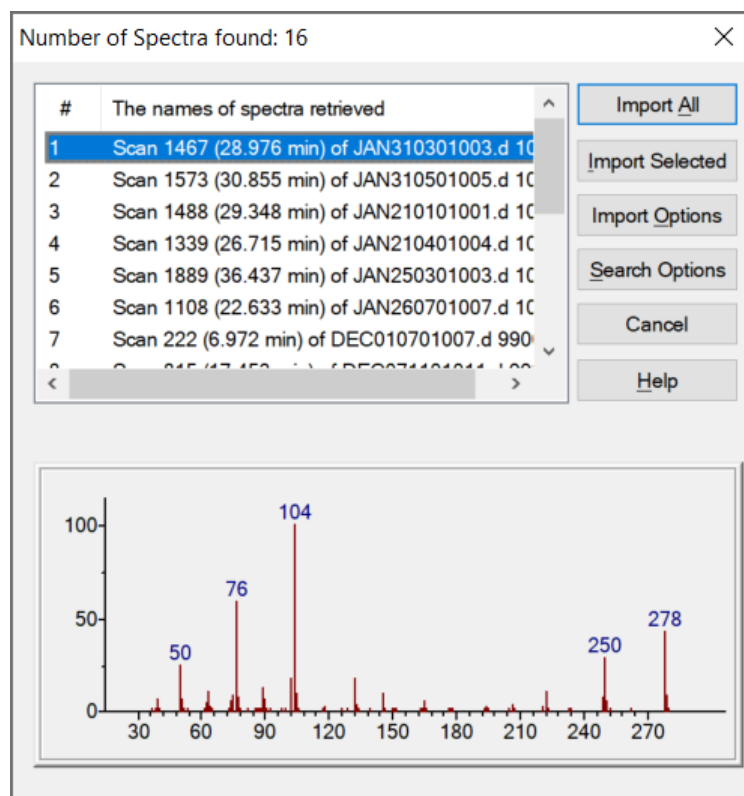


Figure 23. Spectrum Import dialog box with a single spectrum from a multiple spectrum file highlighted and displayed in a bar-graph format.



**Figure 24. Spectral Import Options. Imported spectra integer  $m/z$  and no mass defect correction.**

The two fields on the upper left of the **Spectral Import Options** dialog box are used to make **mass defect** corrections. The difference in the integer nominal mass and the exact mass of an isotope is the **mass defect**. The most significant effect of mass defect is seen on the integer mass of hydrocarbons due to the **mass defect** of hydrogen. The integer mass of the  $^1\text{H}$  isotope is 1. The exact mass of this isotope is 1.00725. The  $\text{C}_{50}\text{H}_{100}$  ion has a nominal mass (the sum of integer mass of the most abundant naturally occurring isotope of each element times the number of atoms of those elements) of 700,  $(50 \times 12) + (100 \times 1)$ . The integer value for  $\text{C}_{50}\text{H}_{100}$  based on a calculation of the exact mass of each of these elements is 701  $(12.000000 \times 50) + (100 \times 1.007825) = 700.7825$ . The **mass defect** for  $\text{C}_{50}\text{H}_{100}$  is 0.7825, which is equal to about 100 millimass units per 100 mass units. The **MS Search Program** deals with only integer mass values when searching the **NIST 20 MS Library**. Spectra are imported with integer  $m/z$  values;

therefore, spectra recorded with observed values of  $m/z$ , which are not integers and have not been corrected for **mass defect**, must have a **mass defect** correction applied to be properly searched against the **NIST 20 MS Library** or another library that contains only integer  $m/z$  values.

The mass defect factor can be entered in the box "Multiply  $m/z$  in imported spectra by". The factor for hydrocarbons based on a general formula of  $-(\text{CH}_2)_n$  is 0.99888. When dealing with spectra that have significant mass contribution from multiple atoms of chlorine or bromine (both of which have a significant negative mass defect), a different factor should be calculated. In this case, it may be desirable to use the field preceded by "Add following term [value] to all  $m/z$  values before rounding".

The "Adding spectra to Spec List" box allows the permanent specification of the action when importing spectra. Choosing Prepend will always add to the current Spec List, Overwrite will always replace the current Spec List. Ask was the only option prior to version 2.2 of the search program.

Intensity threshold allows you to remove low intensity peaks from the spectrum as it is being imported. This can be done in terms of relative peaks – for example 0.1% would keep peaks to 1 part in 1000 of the base peak. Absolute will simply remove data below a user specified threshold.

Spectra without precursor ion  $m/z$  values are assumed to be EI data by default, but can be changed to be In-source/EI with accurate ion  $m/z$  spectra.

RI data in input spectra are not normally characterized by type of column. If you want to specify a particular column type for your data you may do so here.

Finally, Synonyms may be included or excluded



## Library Search Options

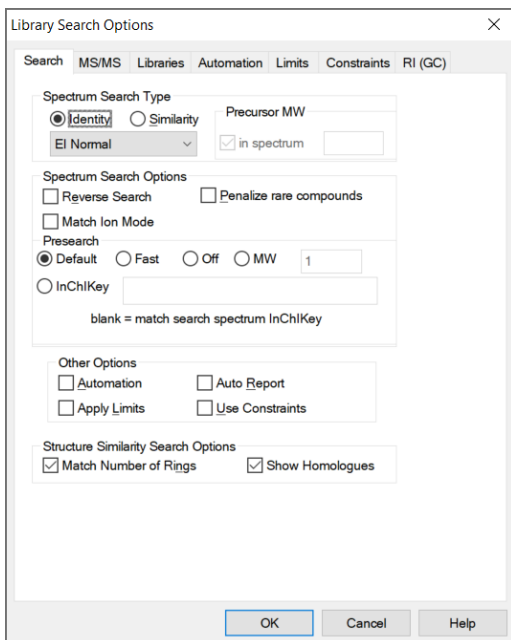


Figure 25. Library Search Options Search tab for an “Identity” Normal search.

Selecting **Library Search Options** on the **Options** menu from any tab or selecting the **Library Search Options** button on the button bar of the **Lib. Search** tab will display a multiple-tab **Library Search Options** dialog box (Figure 25).

A number of different types of spectral comparison routines for choosing the best matching spectra are available. The routine to be used is selected from the **Library Search Options** dialog box’s **Search** tab.

The two search types are **Identity** (with four options) and **Similarity** (with five options). These different searches and their options are explained in the **Help Screens** (topic Spectrum Search Type).

Generally, if it is believed that there is a spectrum of the unknown compound whose spectrum is being searched in one or more of the libraries being searched, an **Identity Search** is best used. If it is believed that there is no spectrum of the unknown compound for a reason such as obtaining only low **Match** and **Reverse Match Factors** for an **Identity**

**Search**, a **Similarity Search** may yield better results. The EI Hybrid and MS/MS Hybrid Similarity Searches introduced with v.2.3 of MS Search are particularly useful in this respect.<sup>4, 5</sup>

The **Search** tab of the **Library Search Options** dialog box also has a **Structure Similarity Search Options** selection. This has the **Match Number of Rings** *check box*. When selected, the search of compounds that have rings will be constrained to only spectra produced by compounds with a similar structure and the same number of rings as the spectrum being searched. The **Show Homologues** selection is self-explanatory.

Note the **InChIKey presearch**. This is used when the structure associated with the spectrum is known. It is a new powerful that help with multiple libraries. Existing User libraries and third party libraries must be indexed to use this feature. Index is accomplished from the **Tools** menu of the main Menu. Details of the InChIKey presearch are found in the **Help** file.

## Hybrid Search

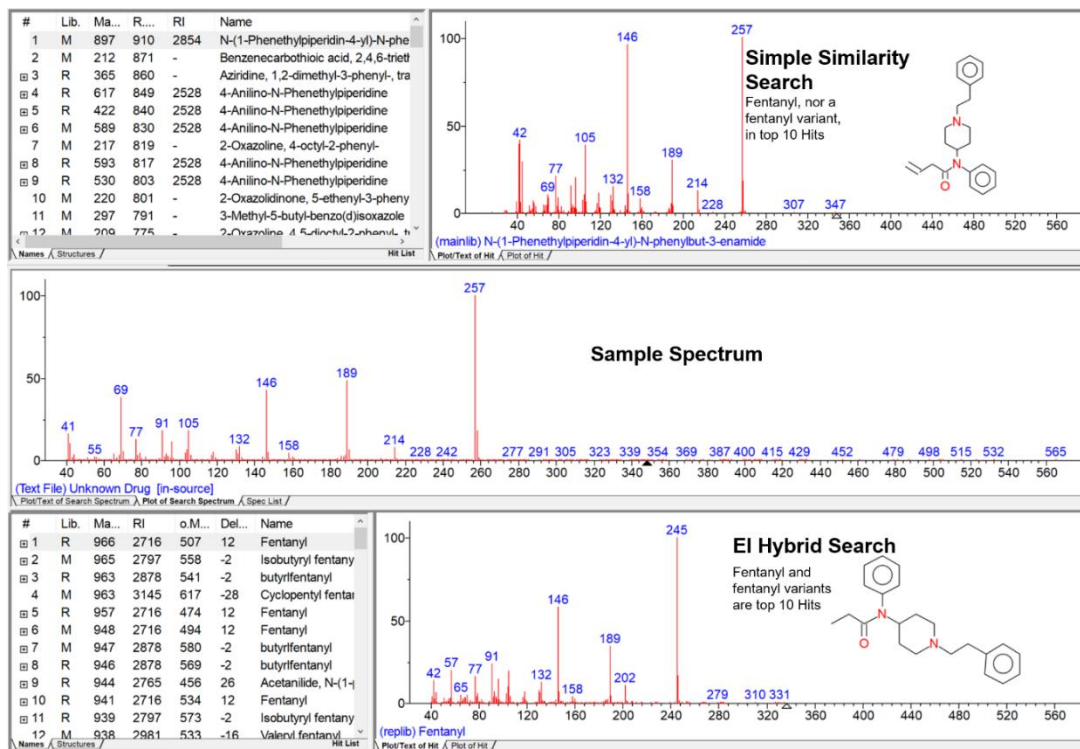
As stated above, an Identity Search is used when it is believed that a spectrum of the unknown compound is in one or more of the searched libraries. A Similarity Search is carried out when there is a high probability that there is not a spectrum of the unknown compound in any of the libraries. The following is an example of how valuable the Hybrid search is when there is no spectrum for an unknown compound in any of the searched libraries. Both the EI Hybrid search and the MS/MS Hybrid search

<sup>4</sup> Arun S. Moorthy, , William E. Wallace, Anthony J. Kearsley,‡ Dmitrii V. Tchekhovskoi, and Stephen E. Stein “Combining Fragment-Ion and Neutral-Loss Matching during Mass Spectral Library Searching: A New General Purpose Algorithm Applicable to Illicit Drug Identification” *Anal. Chem.* **2017**, *89*, 13261–13268.

<sup>5</sup> Ivana Blaženović, Tobias Kind, Michael R. Sa, Jian Ji, Arpana Vaniya, Benjamin Wanczewicz, Bryan S. Roberts, Hrvoje Torbašinović, Tack Lee, Sajjan S. Mehta, Megan R. Showalter, Hosook Song, Jessica Kwok<sup>1</sup>, Dieter Jahn, Jayoung Kim, Oliver Fiehn “Structure Annotation of all Mass Spectra in Untargeted Metabolomics” *Anal. Chem.* **2019**, *91* (3) 2155-2162

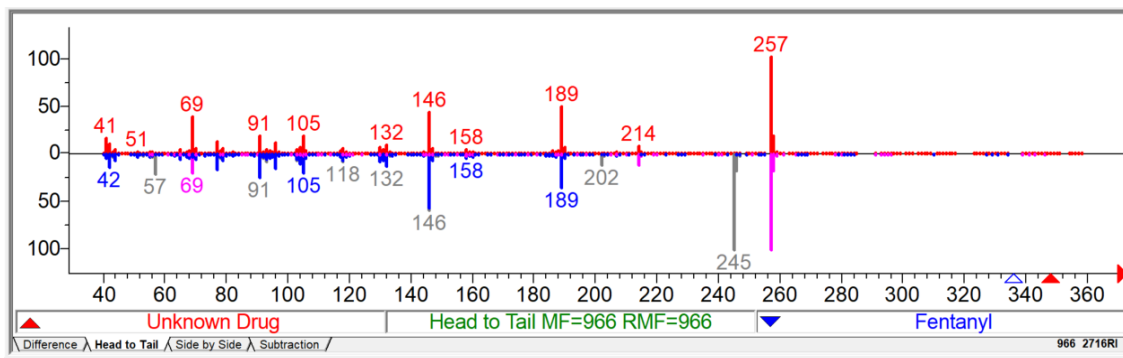
require the use of a precursor ion value. When performing the MS/MS Hybrid search this is requirement is easily met because the  $m/z$  value of the precursor is imported with the spectrum.

In the **EI Hybrid** search, the  $m/z$  value of the molecular ion ( $M^{+}$ ) is required. If the EI mass spectrum does not exhibit a molecular ion peak (of which approximately 20% of the EI spectra in the **NIST/EPA/NIH** EI library do not), then this value will have to be determined using chemical (CI) or field (FI) ionization, by derivation, or as an educated guess. CI was used in the following example for  $m/z$  value of the  $M^{+}$ .



**EI Hybrid Search of Unknow Substance. The top panel shows the results of an EI Simple Similarity Search. Middle Panel is the Raw Spectrum. The bottom Panel Shows the Results of the Hybrid Search.**

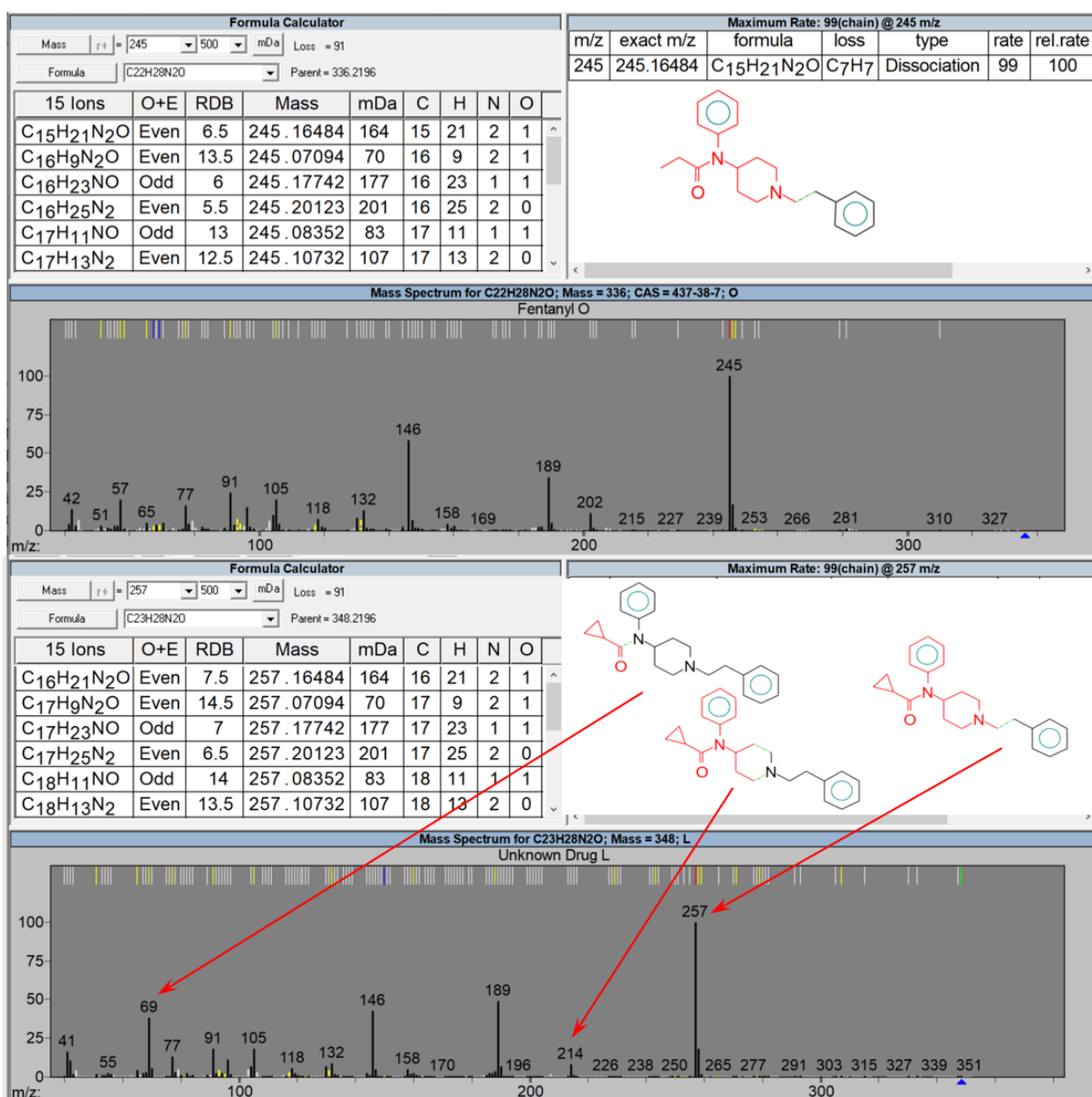
The first Hit of the **EI Simple Similarity Search** is for a fentanyl variant with a relatively good Match Factor and Reverse Match Factor; however, the difference in the  $m/z$  values of the two  $M^{+}$  confirms that there is no identification. There are several spectra in the Hit List of fentanyl variants, but Match Factors (and R Match Factors) are low and the results are defiantly unconvincing.



Comparison of Sample Spectrum (top) and Hybrid Spectrum (bottom) blue and pink peaks. Gray peaks on bottom spectrum belong to Hit and are not part of Hybrid. The difference between them and the and pink peaks represent the DeltaMass

The **EI Hybrid Search** yields much more compelling results. The top 10 Hits are all fentanyl and the **Hybrid Search** Match factors are all at levels above 940. The **oM...** column is the 'Original' Match Factor (obtained before application of the **Hybrid Search** algorithm). An essential column shows **Del...** (**DeltaMass**). This is the difference between the precursor ion of the sample spectrum and the library spectrum. It is used to find peaks 'shifted' by this value.

The top 10 hits consist of Fentanyl or Fentanyl variants, strongly suggesting that the unknown compound is probably a Fentanyl variant that does not have a spectrum in the NIST/EP/NIH Library. Fentanyl has an ethyl moiety (29 Da) attached to a carbonyl carbon atom. A cyclopropyl group (41 Da) is 12 Da greater in mass. If the structure for the cyclopropyl group variant is drawn, associated with the spectrum of the unknown, and then the pair individually sent to MS Interpreter, this tentative identification is confirmed.



The three structures represent unique fragments of the variant and are the rational for the unambiguous identification.

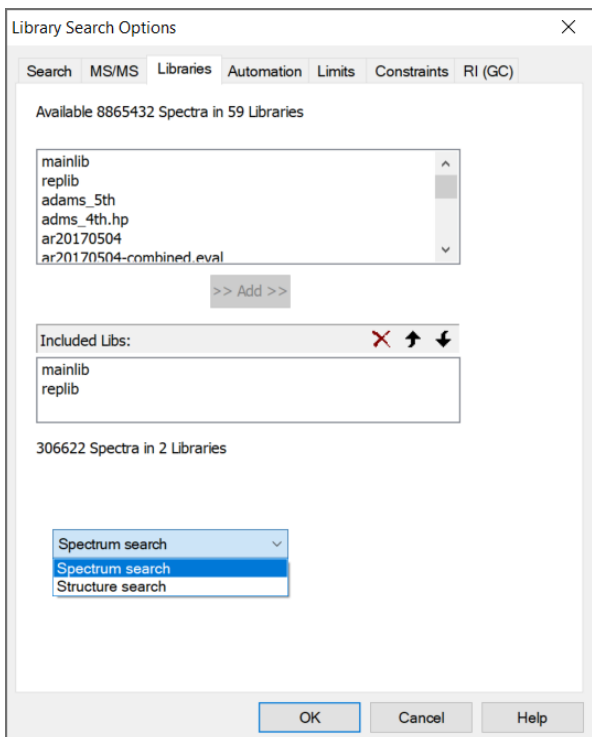


Figure 26. Library Search Options Libraries tab.

accommodated, the less used ones can be moved to another folder. To add libraries to be searched in the **Available Libs** window of the dialog box, highlight them and click on the **Add** button. To exclude libraries, highlight them in the **Included Libs** window and then click on the **X** button above the upper right portion of this window. The order in which libraries are searched can be changed by highlighting the library to be moved in the **Included Libs** window and clicking on the Up or Down arrows in the upper right portion of this window. Each click will move the library one position. The order in which libraries are searched is not important because the **Hit List** has the 100 Highest Match Factor Hits regardless of what library they came from.

Below the **Available Libs** and **Included Libs** windows is a drop-down list box. This box allows for switching between the lists of libraries to be included in **Structure** searches or **Spectra** searches. These two different types of searches can have different sets of libraries.

## Retention Index Data

The retention index search options are generally only available when the search spectrum has a retention index (RI) value in the submitted spectrum. The RI value is taken either from a line in the MSP file of the format

RI=value or from the corresponding text in comment section of the MSP file.

The use of the RI data is controlled by the **RI(GC)** tab in the **Library Search Options** dialog box

## MS/MS

The **MS/MS** tab permits selection of the  $m/z$  tolerances for the precursor and product ions in the **MS/MS** and **In-source HiRes Search**, the option to ignore the precursor ion, and access to a set of options specific to peptides.

## Libraries

The **Libraries** tab (Figure 26) of this dialog box is used to select the libraries to be included in the **Library spectrum Search**. As many 127 libraries can be searched at the same time. This includes the **NIST 20 Main Library (mainlib)** and the **NIST 20 Replicates Library (replib)** and multiple user libraries as well as other commercially available libraries from sources other than NIST that are in the **NIST MS Search** library format. By default, four other **NIST 20** Libraries are also installed: the **nist\_ri**, **nist\_msmsHiRes**, **nist\_msmsLoRes** and **nist\_msms2**. This means that there is space to accommodate an additional 121 user libraries. If more MS libraries need to be

**Error! Reference source not found..**

If the “**Use Retention Indices...**” box is checked, then the program will compare the RI data in the submitted spectrum to the RI value in the library spectrum. The comparison will only be made if the type of column for the search spectrum and the library spectrum are the same.

In general, the data will not have the column type specified so it the specification can be made at the time the data is imported. Alternatively, if the search spectrum column type is not explicitly stated, the program will assume that it is the type specified in the box “**Select RI column Type for Searching or Displaying**”.

The normal mode of operation will restrict the search to data of the same column type. However, since the two non-polar types are somewhat similar, the program has the option to allow the use of the other polar type if the specified type is not present. NOTE: If there is both standard (DB1) and semi-standard (DB5) non-polar data, and you have specified the DB5 data, it will be used. The DB1 data will only be used if a specific library spectrum does not have data for DB5. No such option is available for the polar column data. To select the option to use the alternate non-polar type check the box next to “**Use other non-polar column...**” below the **Select RI Column** box.

When indexing a user library, the option exists to specify the column type. If it is not specified and the column type of the individual RI values for spectra in the library are not specified, then the column type will be set to “User” or unspecified. In this case, you must check the last option to ensure that the RI data in the user library is correctly processed.

A brief note on the values of RI tolerance and RI penalty rate: The RI variation within a set of data taken using basically the same protocols and generally similar matrices will vary by a very small amount, typical standard deviations are in the range of 2-3 RI units. The median variation (that is the median of the uncertainty (median absolute deviation) of the data in the NIST RI library is over 6 RI units. Thus, the data in the NIST library may be at greater variance from your experimental data. For this reason, the recommended RI tolerance is high, and the penalty rate is low.

The use of both non-polar data sets, while expanding the range of RI data must be done with care. Looking at all the data in the collection where there are both DB1 and DB5 data, the DB5 data is about 0.74% greater than the DB1, that is  $RI(DB5) = 1.0074 * RI(DB1)$ . This is a small variance, but at a RI of 2000, it corresponds to a difference of 14.8. In addition, the scatter in the correlation data is large. If the RI tolerance is 10 and the wrong column type is being used since the correct one is not available, there may be too much penalty.



## Automation

When a **Library spectrum Search** or a **Structure Search** is carried out on a spectrum or a structure that is imported from a file or third-party program such as an instrument data system or drawing program, a search can be automatically launched. Automatic searching and/or printing can be turned on or off from the **Search** tab of the **Library Search Options** dialog box by the respective selection or deselecting of the appropriate *check box* (Figure 25 and Figure 28).

The **Automation** tab of the **Library Search Options** dialog box is used in conjunction with these automated searches (Figure 28). *Check box Automatic Search On* (Figure 28) has exactly same functionality as **Automation** *check box* on Figure 25: turns on automatic searching. Selecting or deselecting this *check box* on either tab toggles the *check box* on the other tab.

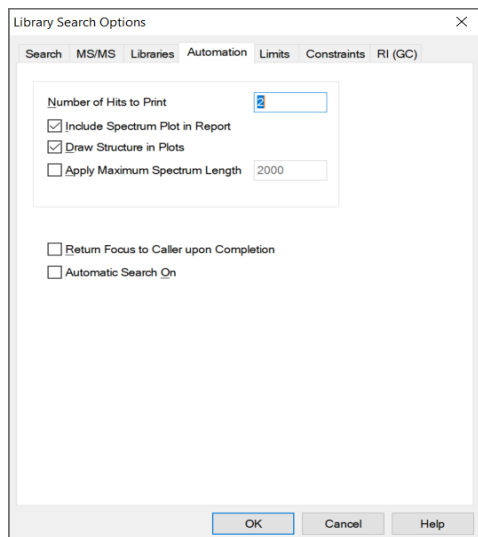


Figure 27. Lib Search Automation tab.

Search results can automatically be printed whether or not the search is automatically launched. *Check box Auto Report* (Figure 25) turns on this feature. “Number of Hits to Print” field is in the **Automation** tab (Figure 28). In addition, the report in the same format can be printed for any displayed search result by selecting **Print Auto Report** from the **File** menu.

The top part of this dialog box pertains to the automated printing of a search result. The format for the Auto Report is a sample spectrum at the top of a page followed by the spectra of a specified number of Hits in the “Number of Hits to Print” numeric entry box (2 by default). If the **Include Spectrum Plot in Report** *check box* is selected, there will be three spectra per page. Page one will have the sample spectrum at the top followed by the first two Hits. Selecting the **Draw Structures in Plots** *check box* will result in the structure associated with a spectrum being printed in the plotted spectrum.

If the **Apply Maximum Spectrum Length** *check box* is selected, a number can be entered for the maximum  $m/z$  value to appear on the spectrum’s abscissa. If peaks are present at higher  $m/z$  values, the abscissa will have a right-pointing arrowhead at its right end. When selected, the **Return Focus to Caller Upon Completion** *check box* will cause a program that brought the **NIST MS Search Program** into focus if already running, or started if not already running, to again be visible after the search.

## Limits

The function of the limits of the starting and ending  $m/z$  values of a **Library spectrum Search** and the lowest value of a mass spectral peak that will be considered in a search can be set. An explanation of the limits that can be used to constrain a search may be found in the Help file by selecting the **Help** button on the **Limits** tab of the **Library Search Options** dialog box.

## Constraints

A **Library spectrum Search** can be constrained by a number of factors such as what elements (and how many atoms of each) must be present in the compounds that produce retrieved spectra, the molecular weight (nominal mass) range of Hits, whether or not the compounds that produce retrieved spectra are in various other (non-mass-spectral) libraries, what  $m/z$  value peaks must be present (along with a range of their relative intensity) in retrieved spectra, what name fragments must be associated with retrieved spectra, what text strings are in the “Comments/Contributor” field of a retrieved spectra, and new to v.2.0g of the **MS Search Program**: an **Exact Mass** Constraint used instead of a molecular weight range. These constraints are explained in the **Help** file, which is displayed by selecting the **Help** button on the **Constraints** tab of the **Library Search Options** dialog box.

## The Search Menu (displayed by Selecting Search on the Menu Bar)

The **Other Search** tab's options are:

Option	Description	Tab View
<b>MW (Nominal Mass)</b>	Finds spectra in selected libraries of compounds having specified (nominal mass) molecular weight (excludes replib).	<b>Other Search</b>
<b>Exact mass</b>	Finds spectra in selected libraries of compounds having an exact mass corresponding to specified accurate mass with a specified variance (excludes replib).	<b>Other Search</b>
<b>MS/MS Precursor <i>m/z</i></b>	Searches for Precursor <i>m/z</i> using precursor-ion mass, formula, <i>m/z</i> values within the uncertainty interval in MS/MS Libraries	<b>Other Search</b>
<b>Formula (Elemental Composition)</b>	Finds spectra in selected libraries of compounds with specified formula (excludes replib).	<b>Other Search</b>
<b>Any Peaks</b>	Finds spectra in selected libraries based on user-specified peaks.	<b>Other Search</b>
<b>Sequential Method</b>	Searches selected libraries using a set of constraints provided by the user.	<b>Other Search</b>
<b>CAS Number</b>	Finds spectra of the compounds that have the entered CAS registry number.	<b>Other Search</b>
<b>ID Number</b>	Finds spectrum (or spectra) in selected library having entered ID number (or range of ID numbers).	<b>Other Search</b>
<b>NIST Number</b>	Retrieves a spectrum with unique NIST Mass Spectral Data Center's Archive ID number (in mainlib and replib only).	<b>Other Search</b>
<b>Name</b>	Allows entering a character string to display spectra beginning with name of a compound.	<b>Names</b>

These options (except for the **Name** search) are also displayed and may be selected from the drop-down list box on the further most left side of the **Other Search** tab (to the right from the **binocular** button). Selecting the **binocular** button starts the currently displayed search in the drop-down list box.

Each **Search** has its own dialog box. When the **MW (Nominal Mass) Search**, **Formula Search**, or **Sequential Method Search** is selected, the dialog box opens to the **Constraints** tab (Figure 29). The **MW (Nominal Mass) Search** and the **Formula Search** dialog boxes have specific entry fields for entering a nominal mass and elemental composition, respectively. Certain obvious constraints are not available (appear grayed in the constraints list and cannot be selected), such as "MW range" and "Exact Mass" in the **MW (Nominal Mass) Search** dialog box and "Element Values" and "Elements Present" in the **Formula Search** dialog box. These three dialog boxes have two tabs. The other tab is the **Options** tab (Figure 30), which allows for the selection of the libraries to be searched. Even though it is present, the **replib** cannot be selected for a **MW (Nominal Mass)**, an **Exact Mass**, or a **Formula Search**. Spectra from the **replib** for these searches automatically appear as replicate spectra in the **Hit List**.

The maximum number of Hits allowed in the **Hit List** can be specified in the **Options** tab of the **Sequential Method Search**. There is also a limit on the maximum number of Hits that can be in a **Hit List** of an **Any Peaks Search** (settable in the **Peaks** tab). The **Lib. Search Hit List** is limited to 100 Hits. There is no limit to the number of Hits that can be in a **MW (Nominal Mass)**, an **Exact Mass**, or a **Formula Search**.

## Search by Any Peaks

Select **Any peaks** from the **Other Search** tab's **Search** menu or the drop-down list box. This will display the **Any Peaks Search** dialog box (Figure 31). Click the " $m/z$ " field (located below the  $m/z$  drop-down list box). The blinking vertical text cursor will now appear in " $m/z$ " field. "Normal" will be displayed in "Type" field. *Type 149 and press <Enter>*. This moves the cursor to the "From" field. *Type 75 and press <Enter>*. This moves the cursor to the "To" field. *Type 100 and press <Enter>*. This will enter the peak with its type and intensity range. The cursor is now in the " $m/z$ " field.

Note that the option of the total number of hits will default to 100, but can be set to a maximum of 1000.

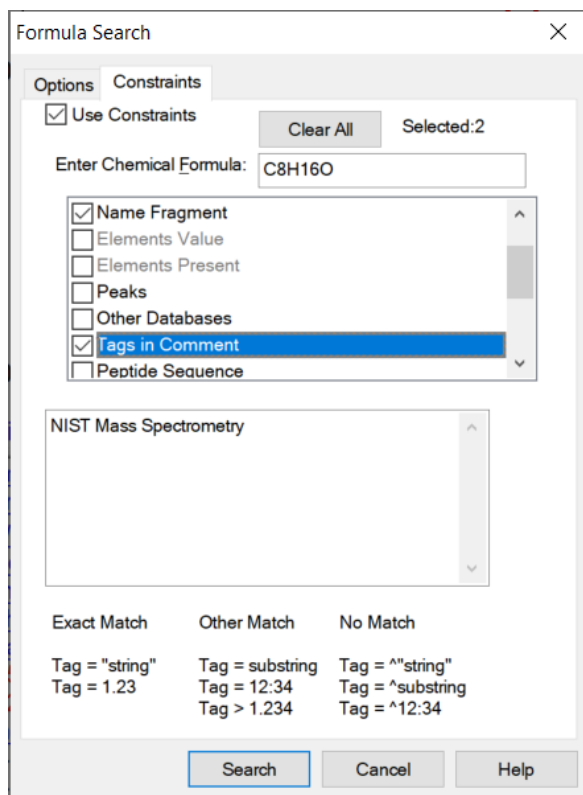


Figure 29. Formula Search dialog box with Use Constraints selected and a Tag in Constraints entered to yield only spectra contributed by Carl Djerassi.

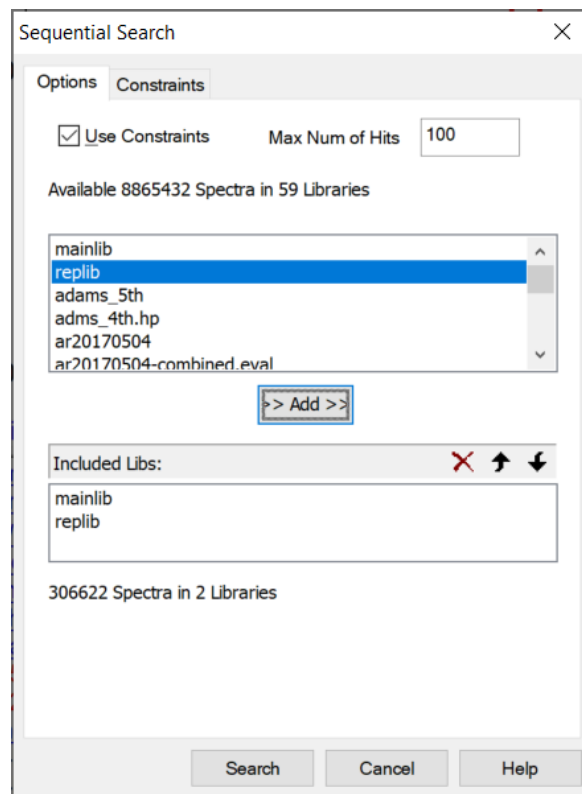


Figure 28. Options tab of the Sequential Method Search dialog box showing that a maximum of 100 Hits will be displayed.



Any Peaks Search

Peak Libraries Constraints

Peak Type m/z Intensity from to

Rank

Accept Delete Clear all


Ty...	m/z	Fr...	To	Hits	Pea...	Hits
N	149	75	100	40...	1	129...
L	0	1	25	12...	2	4080
M	300	1	25	36...	3	62
R	191	4	7	26...	4	1


Polarity for accurate m/z any Peaks required: 4

☐ Use Constraints Max Num of Hits 100

Search Cancel Help

**Figure 30. Peaks tab of the Any Peaks Search dialog box showing the selection of the four different types of peaks and intensity ranges.**

Put the mouse pointer on the  on the right side of the "Type" field and click the left Mouse button. Put the Mouse pointer on "Loss" in the displayed menu and click the left Mouse button again. This will cause the cursor to reappear in the "m/z" field. "Loss" will be displayed in the "Type" field. The first peak entered is an observed peak in the mass spectrum, and it is possible that it is the base peak. The second peak entered represents an ion formed by a loss from the molecular ion. *Type 0 and press <Enter>. Type 1 and press <Enter>. Type 25 and press <Enter>.* The retrieved spectra must have a peak representing the molecular ion peak that exhibits an intensity between 1% and 25% of full scale. This reduces the number of spectra in the **NIST/EPA/NIH Mass Spectral Main** and **Replicates Libraries** that match the entered criteria from 4025 for the first peak to 1302 that have both peaks

The next peak will be a "Maxmass" peak. *Click on the  on the right side of the "Type" field. Click on "Maxmass". Type 300 and press <Enter>. Type 1 and press <Enter>. Type 25 and press <Enter>.* There will only be 17 spectra that contain all three peaks of the designated type and over the designated intensity range.

Note peak types of **Accurate MZ** and **Accurate**

**Loss.** For more information on these two types of peaks, put the Mouse pointer on the Help button in the Ant Peaks Search dialog box and Click the Left Mouse bottom.

## NOTE – USER LIBRARIES ARE NOT SEARCHED USING ANY PEAKS EXACT MASS

The Any Peaks Search can also perform searches on Exact Mass of the fragments. However, this search mode only works on the NIST libraries; it will not function on user libraries even if the user library has been indexed by exact mass. The exact mass indexing of the user libraries done through the Tools/(Re)Index Exact Mass only indexes the precursor or molecular ion based on the formula.

Ty...	m/z	Fr...	To	Pea...	Hits
N	238	1	20	1	298...
E	149.1345 ± 5 m/n	10	80	2	268

**Figure 31. Any Peaks Search dialog box, showing an entry of an accurate mass value with variances. Note the position of the slider bar.**

Figure 32 is the **Any Peaks Search** dialog box set to accept an accurate mass value with variance. It is important to remember that the “Exact Mass” **Any Peaks Search** is only valid for monoisotopic peaks (all atoms of all elements in the ion represented by the peak are the most abundant isotopes). Do not try to enter values for peaks that contain higher mass isotope of chlorine ( $^{37}\text{Cl}$ ), bromine ( $^{80}\text{Br}$ ), or other elements. An accurate mass (**Mass**), an  $m/z$  value (adjusted for electron mass), or an elemental composition (**Formula**) can be entered.

The uncertainty can be in millimass units (mmu) or parts-per-million (ppm). If this field is left blank, an uncertainty of 0.5 less than the last digit entered for the mass is assumed. In this case the uncertainty always is a 0.5 value; i.e., the uncertainty for 120.00 is  $\pm 0.005$  and 120.000 is  $\pm 0.0005$ .

To enter a chemical formula, select **formula** in the drop-down list box above the first entry field (this is possible only if “Exact Mass” type has been selected.) If the “Uncertainty” field is left blank, the accuracy of searched mass is exactly the same as the accuracy of the mass saved in the library. It is very rare the exact mass of two

different elemental compositions is the same; therefore, there is a high probability only the spectra where the peak was interpreted with the same elemental composition will be found. Molecular ion peaks in the mass spectrum will give very good results when used for the **Any Peaks Search Exact Mass** option; however, caution must be taken when using the accurate mass of fragment ions. The exact mass values in the **NIST DB** are calculated from the elemental compositions of ions predicted by **MS Interpreter** where peaks are present in the mass spectrum. More than one exact mass may be saved in a library for a peak in a mass spectrum of a compound. For instance, both CO and C<sub>2</sub>H<sub>4</sub> may be assigned to the same peak with  $m/z$  28. Searching for these two “Formulas” would produce a number of Hits containing both peaks. These Hits correspond to mass spectra where the peak at  $m/z$  28 was ambiguously interpreted. The Any Peaks criteria can be used in the **Sequential Method Search** or as a constraint with a **MW (Nominal Mass)**, **Exact Mass**, **Formula Search** or with a **Spectrum Search**; however, the “Exact Mass” Peak type is not available for these searches. When **Peaks** is used as a Constraint, relative intensity can be specified. In this case, the subsequent peaks have intensities relative to the first peak. Thus, if the first peak has an intensity between 50 and 100% of base, the second peak can be specified to have an intensity of between 40 and 200% of the first peak. This is especially important when using SIM (selected ion monitoring) searches and wanting to understand the rates of false positive identifications. Radio buttons appear at the bottom of the **Peaks Constraints** dialog box to switch back and forth between “Absolute” and “Relative” (1st Peak is Special !).

## Exact Mass Search

**Exact Mass Search** can be used directly to search the NIST libraries. In addition, **Exact Mass** (alternately with nominal mass) can be used as a Constraint in **User Spectra Searches (Lib. Search tab)**, **Any Peaks Search**, and **Sequential Search (Other Search tab)**.

### NOTE – THE EXACT MASS CONSTRAINT IS FOR MOLECULAR OR PRECURSOR ION MASS

Select **Exact Mass** from the drop-down list box on the upper left of the **Other Search** tab. The three-tab dialog box shown in Figure 33 will be displayed. The exact mass value to be searched is entered into the “Search value” field on the **Accurate Precursor Mass Search** tab of the dialog box. The “Type” of the value is selected from the drop-down list box next to the “Search value” field. All the fields in the dialog box are explained in detail by selecting the dialog box’s **Help** button. The **Libraries** tab allows the selection of the libraries to be searched. Like the **MW (Nominal Mass) Search** and the **Formula Search**, the **NIST 20 replib** is not available for searching. The **Constraints** tab allows the **Exact Mass Search** to be constrained by presence in other databases, name fragment, Tags in the “Comments/Contributor” field, elements present, and element values. In addition, the formula or the formula for the neutral loss can be entered. In this case, the program will calculate the exact mass and use the uncertainty (if provided).

The **Exact Mass Search** is not restricted to the **NIST/EPA/NIH Libraries**; however, user or other non-NIST libraries must first be indexed by the exact mass. Select **Tools** from the **Main Menu**

Figure 32. Exact Mass Search's dialog box.

Figure 33. Dialog box to allow for indexing databases according to exact

and then select **(Re)Index Exact Mass**. This will result in a dialog box being displayed (Figure 34). The time to index a library will depend on the size of the library. All **NIST 20 Libraries** are already indexed by the exact mass.

After the desired libraries have been indexed (this has to be done only once), an **Exact Mass Search** or another **Search** using **Exact Mass** as a Constraint can be carried out.

## Search by Name - Browse

The data in a library can be searched or browsed using the incremental name search (the **Names** tab). As a name is entered, the program finds the next name that fits all the text typed to that point. The names search works over not only the primary name but all synonyms. This means that common names for chemicals can be quickly searched. Some common compounds have a very large number of synonyms. Once the name is found highlight the desired name, and the display in the **Plot/Text** and **Plot** Windows will be refreshed with the appropriate display for the compound.

Use the Mouse pointer to place the blinking vertical text cursor in the "Name" field at the top of the **Name Search** tab. To search for the mass spectrum of a plant growth regulator, GIBBERELLIN A9, begin typing the name. Stop after entering each letter. After the letter N has been typed, the 13<sup>th</sup> name in the **Hit List** window is the target. Names may also be pasted from the Windows Clipboard into the "Name" field by simultaneously pressing keys **Shift** and **Insert** or using the RMB menu's **Paste** selection.

NOTE: This feature can be used in a user library to give synonyms such as "LabBook05-Pg6" or more simply "B05-P06", this allows the data to be quickly found by reference to the original experiment.

## Sequential Method

This method allows every spectrum in the selected libraries to be subjected to a created search criterion. The criteria that are set in the Constraints section govern the **Hit List** returned. In this search method it is possible to use not only absolute but also only relative intensities of peaks. This can be an important tool for examining the risk of false positives when using SIM searches.

In the example shown in Figure 35, the peak at 109 is allowed to be present in a range of 1 to 100% of

The Sequential Search dialog box is shown with the 'Constraints' tab selected. The 'Use Constraints' checkbox is checked. A list of search criteria includes MW(Nominal Mass), Exact Mass, Name Fragment, Elements Value, Elements Present, Peaks (which is selected and highlighted in blue), and Other Databases. Below this list, there are input fields for 'Type' (set to 'Normal'), 'm/z' (empty), 'From' (empty), and 'To' (empty). A note says 'Enter mass between 1 and 2000'. There are 'Accept', 'Delete', and 'Clear all' buttons. A table shows the search results:

Type	m/z	From	To
N	109	1	100
N	158	11	35
N	63	20	50

At the bottom, the 'Mode' is set to 'Relative (1st Peak is Special I)' with a radio button. There are 'Search', 'Cancel', and 'Help' buttons at the very bottom.

Figure 34. Sequential Search dialog box showing the use of Any Peaks in a Relative mode.

the base peak. The subsequent peaks (158 and 63) are sought as a fraction of the 109 peak, regardless of the intensity of the 109. If the peak with  $m/z$  109 is 15% of the intensity of the base peak and the peak with  $m/z$  158 is 5% of the intensity of the base peak (i.e., 33% of the 109) it would be accepted. This kind of analysis can give important insights into the risk of false positive matches when using SIM (selected ion monitoring) in GC/MS experiments.

In SIM the entire spectrum is not seen so the relative intensities of the ions observed are the only data that can be extracted from the data file.

NOTE: All the other methods of searching are also present in the Sequential Method. The elements search can also be interesting if the number of compounds in all the libraries that have at least one mercury atom in the molecule needs to be determined. On a modern computer with a large amount of memory these searches do not need to take very long. Typically any sequential search will be faster for subsequent similar searches because much of the data file will be cached in the computer memory.

---

## Basis for Interpretation of the Library Search Results

When an **Identity Search** with an unknown spectrum (**Library spectrum Search**) is performed, the results will be associated with three numbers for each spectrum shown and one value for the search. These three numbers are: 1) a **Match Factor** for the unknown and the library spectrum (direct match of peak  $m/z$  values and relative intensities), 2) a match factor for the unknown and the library spectrum ignoring any peaks in the unknown that are not in the library spectrum (**Reverse Match**), and 3) a **Probability** value (column headings in the **Lib. Search Hit List: Match, R. Match and Prob.**, respectively). The first two numbers are straightforward. Each is derived from a modified cosine of the angle between the spectra (normalized dot product). A perfect match results in a value of 999; spectra with no peaks in common result in a value of 0. As a general guide, 900 or greater is an excellent match; 800–900, a good match; 700–800, a fair match. Less than 600 is a very poor match. However, unknown spectra with many peaks will tend to yield lower Match Factors than similar spectra with fewer peaks. The **Probability** value and the value for the search (**InLib**, located in the lower right corner of the **Hit List** window and on the **Title Bar** of the **Lib. Search** tab) require more explanation.

The **Probability** value for a Hit is derived if the compound is represented by a spectrum in the libraries searched. It employs only the differences between adjacent Hits in the **Hit List** to get the relative probability that any Hit in the **Hit List** is correct. This value is derived from an analysis of the results of searching the **NIST/EPA/NIH Main Library** with a set of replicate spectra (given in the **Replicates Library**). The relative probability of each of the Hits requires only the difference values because the total probability of the compound being in the searched libraries is assumed to be one. The other factor (**InLib**) is a measure of the probability of the compound being in the searched libraries. This value is also derived from the same set of replicate spectral searches. In this case, the correct compound was ignored in the **Hit List**; and the difference between the **Hit List**, with and without the correct compound, is parameterized. The parameters are the maximum value of the match and the largest single difference among the top 20 Hits. If the first Hit has a high Match Factor (>900) and the next Hit has a Match Factor of 800 or less, the probability of the compound being correctly identified is very large and the probability of the compound being in the searched library is large (the number of “Hits vs. their Match Factors” is displayed in the **Hit Histogram** window located just above the **Hit List**, see Figure 1).

Like all statistical results, these probability calculations rely on the data sampled. For example, if a compound that has very few similar mass spectra will be identified in a more definitive way (a higher **Probability** and **InLib** value). Using examples from the **Replicates Library** and searching the **Main Library** using a compound like 'folpet', will result in a high probability for the first Hit and a high **InLib** value. In contrast, if replicates of 'cyclohexanol' are searched, the results will be much lower **Probabilities** and **InLib** value; and, in some cases, the best **Match** is not even the correct compound. This reflects the fact that there are very few compounds that have mass spectra similar to 'folpet'; but there are a number of compounds that have very similar mass spectra to 'cyclohexanol', and the ability of any search system to distinguish between these similar spectra is limited. In many cases, the best that the search can do is to provide a class of compounds that have similar mass spectra and, usually, similar structures.

The values of the **InLib** parameter are meant as guideposts. Generally, any positive value is acceptable. Values greater than approximately 300 usually mean that the spectrum is nearly unique. Negative values below 200 are generally a warning that the spectrum is not identified. Note that negative values will occur when there are many compounds with similar spectra. In these cases, the difference between the Match Factors for different spectra is very small, and the search cannot be assured of providing the correct unique answer. Usually in these cases, especially when Match Factors are high, it will provide very good guidance on the structure of the molecule.

For a complete discussion of the methods used in assessing the probabilities, see Stein, S.E. “Estimating Probabilities of Correct Identification from Results of Mass Spectral Library Searches” *J. Am. Soc. Mass Spectrom.* **1994**, 5, 316–323.

## Substructure Information

The **Substructure Information** utility is used in the evaluation of a **Hit List** using the **NIST/EPA/NIH Mass Spectral Library** because each spectrum is indexed by the substructures that make up the structure of the compound that resulted in the spectrum. Substructures are functional groups like carbonyls, methoxys, methyl ester moieties, etc.; the number of rings and double bonds; and heteroatoms.

The **Substructure Information** is most useful when the **Hit List** is the result of one of the four **Similarity Searches** (**El Hybrid**, **El Simple**, **Neutral Loss**, **MS/MS in El**, or **MS/MS Hybrid**). This utility is called from the **Main Menu Bar's Tools** menu.

**Similarity Searches** are carried out when it is believed that a spectrum of the unknown compound is not in any of the searched libraries. Spectra similar to the searched spectrum make up the **Hit List**. There is no **Prob** column in the **Hit List** and an **InLib** value is not determined. Not only is the utility useful in determining the structure of an unknown based on its EI mass spectrum, it has proved to be very informative in the analysis of collision-induced dissociation (CID) spectra obtained from ions produced by soft ionization LC/MS techniques such as APCI, APPI, and electrospray. This use is described in more detail below.

The **Substructure Information** dialog box (Figure 36) has the name of the spectrum that was used to create the **Hit List** displayed at the top. Just below the "Name of Unknown" field in this dialog box is the "Chlorine/Bromine information" field, which gives the probabilities and numbers of atoms of chlorine and/or bromine present (useful regardless of the method by which the spectrum was obtained). The result of molecular weight estimation from the **Hit List** is presented on the right side of the dialog box

Substructure Information

Name of Unknown  
Scan 372 (6.199 min): 50P0830D.D\data.ms

Chlorine/Bromine information  
Cl=3, Br=0 Probability=82%  
Probability of presence of Cl=97%, of Br=0%

MW Prob. (%)  
132 40  
117 10

Substructural information

Prob. (%)	Present	Prob. (%)	Absent
99	halogen	99	OH
99	Cl	99	4+brid
99	ClBr	99	Ar-O-F
98	RCI	99	Ar-O
96	nocyc	99	Ar-C
94	NoAr	99	ArO
90	Ctert	99	Si(CH

Set of Substructures in use

#	Substruct
1	OH
2	CO2H
3	ArOH
4	ROH
5	SH
6	?OH
7	SiH3

Customize

OK Print Help

Figure 35. Substructure Information dialog box.

(useful for EI spectra, especially when no molecular ion peak is observed). There are two windows below this in an area labeled **Substructure information: Prob. Present** and **Prob. Absent**. These two windows contain a list of abbreviations for the substructures preceded by a number, which is the percent probability for the presence or absence of the substructure. If one of the abbreviations is not understood, highlight it by using the Mouse and a detailed explanation appears at the bottom of the **Substructural information** area.

Just below the molecular weight information is the **Set of Substructures in use** window with a list box and a **Customize** button. The list box displays all the substructures in the current set.



The **Customize** button, when selected, displays the **Customized set of Structures** dialog box (Figure 37). This dialog box is used to create a subset of the substructures to be identified. The list of all possible substructures is displayed in the **Full set of structures** window. The list of substructures included in the

Substructure Information

Name of Unknown  
Toluene

Chlorine/Bromine information  
Cl=0, Br=0 Probability=99%  
Probability of presence of Cl=0%, of Br=0%

Substructural information

Prob. (%)	Pres. ^	Prob. (%)	Abse ^
99	CH2/	99	OH
98	RDB	99	hetcy
90	HCU	99	Ar-O-
90	HC	99	-O-
72	ph	99	RDB
72	AR	99	CCH
69	Ph-C v	99	ArOF v

Set of Substructures in use

#	Substruct. ^
1	OH
2	CO2H
3	ArOH
4	ROH
5	SH
6	?OH
7	SiH3 v

Customize

OK Print Help

Figure 36. Customized Set of Structures dialog box.

the **OK** button. This dialog box, as well as the **Substructure Information** dialog box, has a **Help** button, which displays a context-sensitive Help screen.

The algorithms used in the substructure identification are described in the Help screen. They are based on developments at NIST (Stein, S.E. "Chemical Substructure Identification by Mass Spectral Library Searching" *J. Am. Soc. Mass Spectrom.* **1995**, 6, 644–655).

present file is shown in the **Customized set** window. The first time a customized list is created, both windows will contain the same list. Substructures may be omitted from the **Customized set** window or added from the **Full set of structures** window. Either action occurs on the structures highlighted by Mouse action in the ws by selecting the appropriate button **← Omit ←** or **→ Add →**. The usual Windows conventions for selecting multiple items apply. After selecting the first substructure of the list, hold down the <Shift> key while selecting the last substructure. This will select all substructures between the first and last. Multiple noncontiguous substructures can be selected by holding down the <Ctrl> key while selecting each desired substructure.

The **Save** and **Open** buttons allow the saving and retrieving of customized substructure sets. A temporary set is created by making modifications to the **Customized set** window and selecting

---

## Example of Use of Substructure Information in Making Determination about Unknown Mass Spectrum obtained using CID

The mass spectrum in Figure 38 of a compound with the structure shown inserted in the spectrum was obtained using collision-induced dissociation (CID) on a protonated molecule with  $m/z$  248. An **EI in MS/MS Similarity Search** was performed against the **NIST 20 mainlib** with  $m/z$  248 designated as being the precursor ion. This type of search is provided to allow for the use of the vast amount of information in the EI Library with spectra that are produced by LC/MS CID. The **EI in MS/MS Similarly Search** has been optimized for this purpose. The resulting **Hit List** was evaluated using the **Substructure Identification** utility. Based on this evaluation, it was determined that there was a 90% probability that the compound contained one or more atoms of oxygen; a 90% probability that there was a carbonyl group; an 89% probability that there was a heterocyclic ring; and an 88% probability that there was an aromatic ring; but it also revealed a 99% probability that there was not a phenyl carbonyl or a methoxy phenyl group, and a 90% probability that there was not a heterocyclic ring that contained oxygen. Remember, it is just as important to know what is absent as it is to know what is present. The evaluation revealed that there was an 89% probability that there were one or more atoms of nitrogen and an 86% probability that there was a heterocyclic ring with an atom of nitrogen. Continued evaluation of the **Prob. Present** and **Prob. Absent** windows in the **Substructure Identification** dialog box revealed other possible substructures to be present and eliminated others by their high-absence probabilities.

This particular use of the **MS/MS in EI Similarity Search** in conjunction with the **Substructure Identification** has not been extensively tested by NIST; however, anecdotal evidence has shown this to be an effective combination and should be considered when trying to determine the structure of an unknown using CID data. In the example cited above, if the unknown was believed to be a metabolite of some known substances, the information gained from this technique could well result in identification.



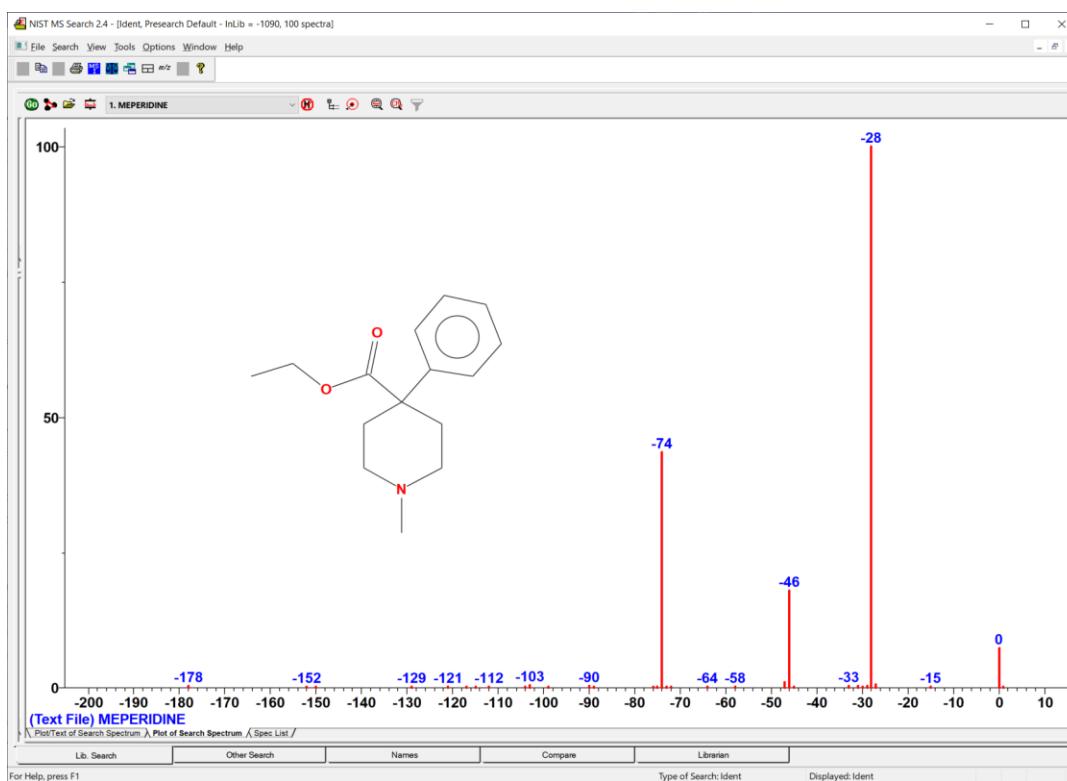


Figure 37. Mass spectrum obtained by CID of ion with  $m/z$  248 and structure.

---

## Use with Instrument Data Systems

The features described in the **AUTOMATION** section of this manual were implemented to aid the use of the **NIST MS Search Program** with proprietary mass spectral data systems.

### Location of the NIST MS Search Program

It may be necessary to determine the location of the **NIST MS Search Program** files and the location of the **NIST MS Search** working directory. These may be found in the **[NISTMS]** section of the **WIN.INI** file located in the %windir% directory. Below is a typical example of this section:

[NISTMS]

```
Path32=C:\NIST20\MSSEARCH\  
WorkDir32=C:\NIST20\MSSEARCH\  
Amdis32Path=C:\NIST20\AMDIS32\  
AmdisMSPath=C:\NIST20\AMDIS32\
```

The string beyond "Path32=" is the path to **MS Search Program's** folder (location of the executable files); the string beyond "WorkDir32=" is the path to the **MS Search Program's** working directory (location of the history, Hits, etc.). Two other items refer to AMDIS and its connection to the **MS Search Program**.

Cases have been reported where the connection between MS Search and a third-party data analysis system was rendered nonfunctional when this [NISTMS] area of the win.ini file referred to more than one location for various version of the NIST MS Search Program.

### Implementation

If the mass spectrometry data system being used is capable of exporting spectra to the NIST format (described in the **USER LIBRARYIES** section of this manual) and a Windows program can be executed from within the data system, this feature is available. Implementation of this feature has been provided by some manufacturers on their data systems.

The command string used to start or bring into focus the **NIST MS Search Program** for Windows is:

```
C:\NIST20\MSSEARCH\NISTMS$.EXE<space>/instrument
```

assuming the program is in the **C:\NIST20\MSSEARCH** directory; if it is in some other directory, make the appropriate substitution.

Two locator files are required to use the automation features:

1. The first locator file must be in the **MS Search Program's** working directory, have the name **AUTOIMP.MSD**, and contain the text string describing the name and location of the second locator file. For example:

```
C:\MYMSDS\FILESPEC.FIL
```

where: C:\MYMSDS is the folder where the file, FILSPEC.FIL, resides.

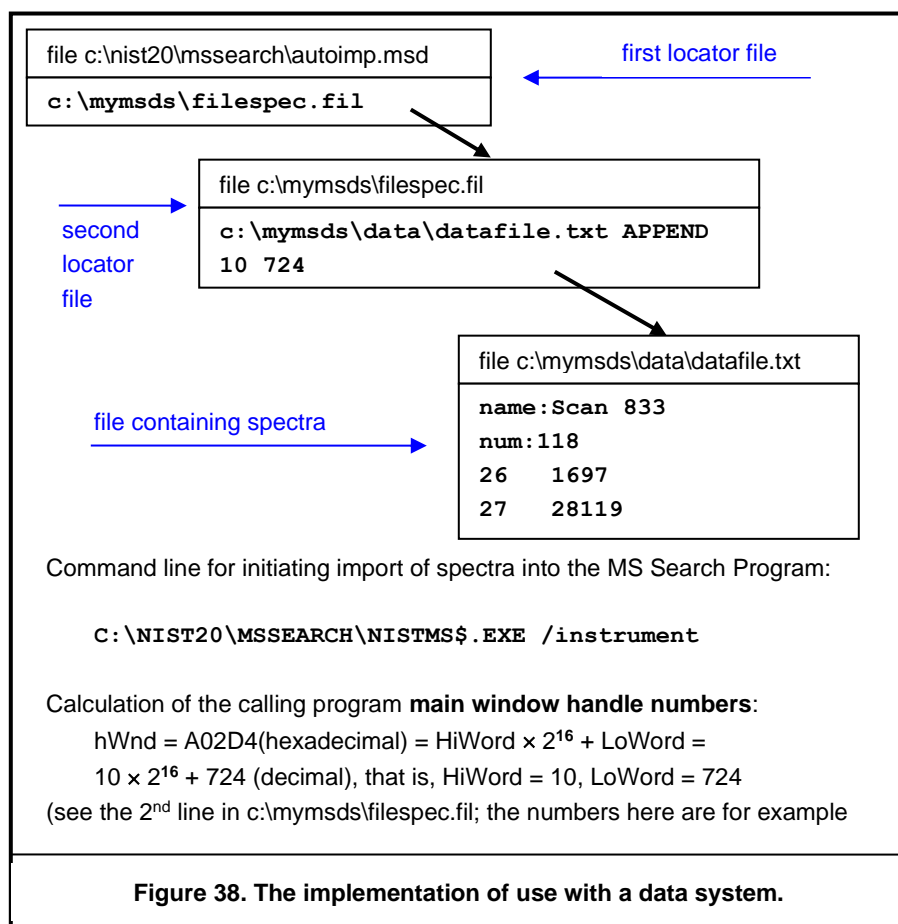
This file can be created with any text editor.

2. The second locator file is created each time a spectrum or spectra is/are saved to a separate file to be imported by the **MS Search Program**. This second locator file contains the text string that describes the name and location of the text file containing the spectrum (or spectra) in the NIST Text file format described later. It is possible to state whether the spectra currently in the **NIST MS Search Program's Spec List** window are to be added to or replaced by the automatic import. An example of the contents of the second locator file is:

C:\MYMSDS\DATA\DATAFILE.TXT<space>OVERWRITE or APPEND

where: C:\MYMSDS\DATA\DATAFILE.TXT is the full pathname of the file containing the spectra, and one of the words OVERWRITE or APPEND is the instruction as to how the spectra are added to the **Spec List Window**. To enable switching the focus back from the **MS Search Program**, the second line containing **main window handle numbers** of the data system application may be added (see Figure 39 for details).

This second locator file is deleted each time the **NIST MS Search Program** is brought into focus and spectra are imported. Figure 39 illustrates the implementation. The upper part of each box contains a file name; the bottom part is the contents of the file. Black arrows show logical connections between the files.



NIST MS Search (nistms\$.exe) Command Line Options	
Option	How it works
/INSTRUMENT	Normal search. Imports and optionally searches spectra (see Figure 39 for details on the implementation), activates <b>MS Search</b> Window, activates <b>switch back</b> button.
/INSTRUMENT /PAR=2	Background search. Imports and optionally searches spectra (see Figure 39 for details on the implementation), does not activate <b>MS Search</b> Window (in fact, it is disabled during the search), writes <b>Hit Lists</b> into file SRCRESLT.TXT; when the search is complete creates file SRCREADY.TXT as a signal that SRCRESLT.TXT is ready. Both files are created in the <b>MS Search</b> work directory. Number of Hits is determined by the "Number of Hits to Print" in the <b>Automation</b> tab (Figure 28).
/PAR=4	Start logging. Appends <b>Hit Lists</b> of <b>Spectrum</b> and <b>Structure Searches</b> to the file NISTLOG.TXT located in the <b>MS Search</b> work directory. Number of Hits is determined by the "Number of Hits to Print" in the <b>Automation</b> tab (Figure 28).
/PAR=8	End logging. Stops appending <b>Hit Lists</b> to the file NISTLOG.TXT.
/MOL="<full pathname>"	Imports chemical structure from a Molfile or SDfile. If SDfile contains more than one Molfile, only the first one will be imported. See section <b>USE WITH THIRD-PARTY DRAWING PROGRAMS</b> for details.

SEE APPENDIX 1 for instructions on the Creation of AUTOIMP.MSD.

---

## Use with Third-Party Drawing Programs

There is a **Structure Search** command. If a structure is in the **Spec List** window of the **Lib. Search** tab or a structure is associated with a spectrum, this structure can be searched to retrieve spectra of all compounds with similar structures by selecting the **Structure Search** button (second button from the left) with the structure highlighted in the **Spec List** window or by use of the Right Mouse button (RMB) menu with the Mouse pointer on the spectrum containing a structure.



When the **NIST MS Search Program** saves a structure to the Windows Clipboard using the Right Mouse button menu's command **Copy Structure to Clipboard**, the structure is copied to the Windows Clipboard and a file in Molfile format named Clipboard.MOL is written to the **NIST MS Search Program's** directory (by default, C:\NIST20\MSSearch). Other programs can either use the structure as it appears on the Windows Clipboard or this file for structure importation.

Structures saved to the Windows Clipboard may be accepted in the **MS Search Program's Librarian Spectrum Information** dialog box if its **Clipboard Struct** button is not grayed. Some drawing programs need to be told how to save a structure in the Windows Clipboard. For example, in ISIS/Draw (MDL Information Systems, Inc.), it is necessary to select **Copy Mol/Rxnfile to the Clipboard** check box in **Options/Settings/General** tab of that program.

If an application is associated with files that have an .MOL extension, then the structure may be sent to that application by selecting in the RMB menu **Send To** and **Default Structure Editor**.

If a spectrum containing a chemical structure is selected, the RMB menu's command **Send to** will result in a **Send to** menu that can have up to three additional structure editor items. These **Send to** menu items are a function of the first three lines in a text file **AUTOIMP.STR** located in the **NIST MS Search Program's** working directory (the WorkDir32 item in [NISTMS] section of the WIN.INI file). Each line should contain three double-quoted strings separated by spaces:

- 1) A program name as it appears in the **Send to** menu
- 2) Program executable and path
- 3) Program startup parameters. %1 will be replaced by the mol file name.

The following is an example of the contents of an AUTOIMP.STR file:

```
"ChemSite" "C:\MolSuite\ChemSite\ChemSite.exe" "%1"  
"ISIS Draw" "D:\ISIS Draw\IDraw32.exe" "%1"  
"ACD/ChemSketch" "C:\acd800\chemsk.exe" "%1"
```

The last line of the above example results in ACD lab's ChemSketch starting with the contents of clipboard.mol being imported when ChemSketch opens.

When a program that is to be interactive with the **NIST MS Search Program** is installed, it must UPDATE or CREATE the **AUTOIMP.STR** file, or the **AUTOIMP.STR** file must be edited by the user before the third-party program will appear on and function from the **Send to** menu.

### SEE APPENDIX 1 for instructions on the creation of AUTOIMP.STR.

A third-party program can call the **NIST MS Search Program** by issuing the command:

```
C:\NIST20\MSSEARCH\NISTMS$.EXE /MOL=<molfile>
```

assuming the **NIST MS Search Program** is in the C:\NIST20\MSSEARCH directory;  
if it is in some other directory, make the appropriate substitution (see **Location of the NIST MS Search Program** above)

where <molfile> is a full pathname of the MOL file. If there is a space in the name (e.g., example c:\program files\a.mol), the name should be double quoted:

```
C:\NIST20\MSSEARCH\NISTMS$.EXE /MOL="c:\program files\a.mol"
```

A third-party program can obtain the path to NISTMS\$.EXE from Path32 in the [NISTMS] section of WIN.INI (see **Location of the NIST MS Search Program** in the **USE WITH INSTRUMENT DATA SYSTEMS** section above).

## User Libraries

The **NIST MS Search Program** can create user libraries to search against. Up to 127 MS libraries, including the libraries from **NIST 20 MS Library** are allowed. The user libraries may be searched in conjunction with the **NIST/EPA/NIH Mass Spectral Main** and/or **Replicates Libraries**, or they may be searched separately. The **Program** allows for the searching of combinations of libraries in any order. The inclusion of libraries and their order of search can be set from the **Libraries** or **Options** tab of the various **Search** dialog boxes (see Figure 26 and Figure 30 as examples).

User libraries are maintained as subdirectories under the **MS Search Program's** working directory. The name given to the user library is the name of the subdirectory. This subdirectory will contain a number of library files and indices constituting the user library. The spectra and compound information are put in these files through the **Librarian** tab or by using **Lib2NIST** library utility.

The spectra and compound information can originate in a text file in the NIST Text file format described in the **NIST TEXT FORMAT of INDIVIDUAL SPECTRA** section of this manual or directly imported from AMDIS. Spectra can be edited using with the **Edit Spectrum** button in the **Librarian** tab **Tool Bar** (Figure 41). This will call up the **Spectrum Information** dialog box (Figure 40). This dialog box can be

m/z	Abund.	Annotation
14	2	
15	38	
16	1	
26	2	
27	9	
28	37	
29	10	
30	3	
31	1	

Accept ☐ HiRes Spectrum Peaks 92

100  
0  
20 40 60 80 100 120 140 160 180 200

55 67 82 109 194

Additional Info

Comments  
NIST Mass Spectrometry Data Center, 1998.

To Clipboard  
From Clipboard  
Get MolFile  
Get Struct  
Structure Editor

MAINLIB via CAS number

Add to Library Add to List Replace Cancel Help

**Figure 39. Spectrum Information dialog box displayed in the Librarian tab view.**

used to add structures to user spectra from the clipboard, add identification information such as CAS number or synonyms and to add comments. In addition, the RI value for the spectrum can be either added or edited.

In addition, a new spectrum can be entered using the **New spectrum** button in the **Librarian** tab **Tool Bar** (Figure 41).

The User libraries are easy to create and maintain. It may be desirable to create them using specific sets of spectra from the **NIST/EPA/NIH Main** or **Replicates Library**. Any spectrum in the **Spec List** can be added to a user library. Any spectrum in the **Hit List** of the **Lib. Search** tab, **Other Search** tab, or **Names** tab, a **Plot** window, **Text Info** window, or **Subtraction** window in the **Compare** window of the **Lib. Search** tab or the **Compare** tab can be copied to the **Spec List** window by selecting **Spec List** on the **Send to** menu from the Right Mouse button menu. The size of a user library is limited to 65,535 spectra.

The **MS Search Program** can read user-submitted spectra in the JCAMP, U.S. EPA, or NIST Text file formats as well as formats of files submitted by various instrument manufacturers' mass spectrometry data systems. The manufacturer of a specific mass spectrometer's data system should be checked with to see if there is file compatibility.

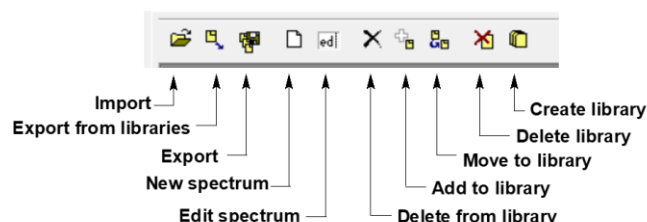


Figure 40. Librarian Tool Bar.

Instructions on building user libraries are included in the **Librarian** Help screens. They also include the information necessary to add structures to user library spectra.

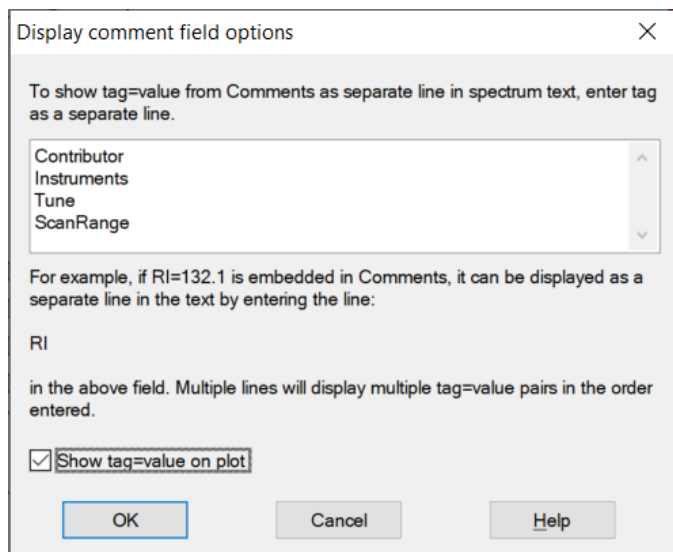


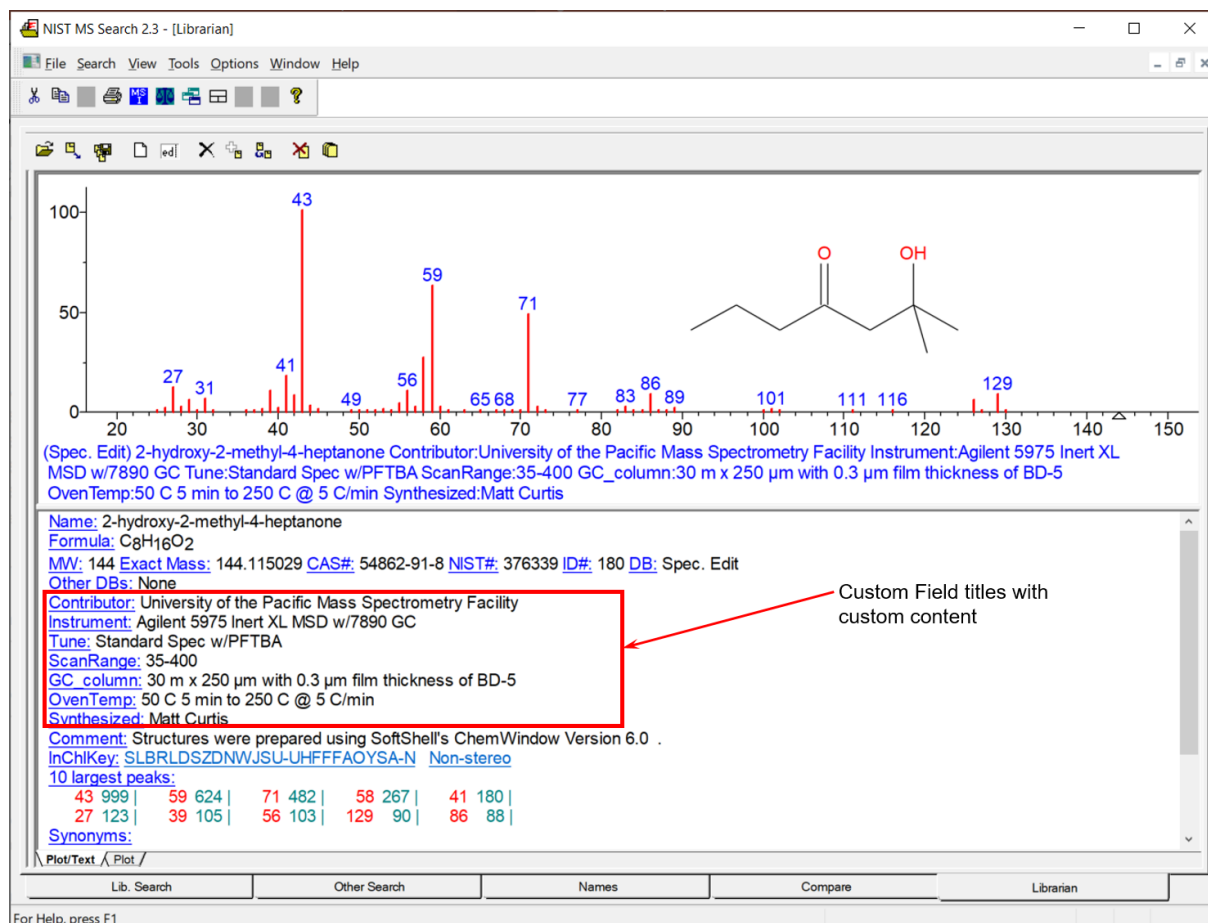
Figure 41. Display comment field options dialog box.

The "Tags" in the "Comments" field of a user library spectrum can be used. A *Tag* is the name of a Field heading that will be displayed with the *Text Information* of a user library or **Spec List** spectrum and can optionally appear in the line below the *Plot* of the spectrum. Text is designated as a *Tag* in the "Comments" field of a user library or **Spec List** spectrum if it consists of the Field title (a line of characters without spaces) followed by an = (equals sign). Following the = is the contents of Field. This can be a line of characters without spaces, a number, or a text string enclosed in quotation marks (" "). For example, the following may be entered in the "Comments" field of the **Spectrum Information** dialog box of a **Spec List** or a user library spectrum:

Structures were prepared using SoftShell's ChemWindow Version 6.0 Contributor="University of the Pacific Mass Spectrometry Facility" Instrument="Agilent 5975 Inert XL MSD w/7890 GC" Tune="Standard Spec w/PFTBA" ScanRange=35-400 GC\_column="30 m x 250 µm with 0.3 µm film thickness of BD-5" OvenTemp="50 C 5 min to 250 C @ 5 C/min" Synthesized="Matt Curtis".

As shown in Figure 43, the information preceding each equals sign (=) will be displayed as a Field title in the **Text Info** window of the spectrum and can optionally be displayed with the **Plot** window of the spectrum. The information following the equals sign (=) will appear following the Field header. Entry of Field contents (*string*) must conform to the rules above. For this display to appear, the Field titles must be registered in the **Display comment field options** dialog box shown in Figure 42 (called from the **Comment field options** selection of the **Options** menu. When the *Tags* have been entered in the **Comment field options** dialog box, they will not be displayed in the “Comments” field of the **Text Info** window.

When a search is performed of a user library that contains *Tags* (Field titles), it is possible to constrain that search based on the *Tag* (Field title) and the partial or whole contents of the *Tag* (Field title). More specific details on the use of *Tags* as Constraints are in **MS Search Program’s** Help file.



**Figure 42. Example of custom field titles with custom content added to a user library spectrum through the use of *Tags*.**



---

## NIST Text Format of Individual Spectra

The NIST Text file format is simple. A file can contain as many spectra as wanted. Each spectrum must start with the Field title "Name:". There must be something in this field in order for the spectrum to be imported (i.e., the data file name and the scan number of the spectrum's source). The next required Field title is "Num Peaks:". The contents of this field must be the number of mass/intensity pairs that make up the spectrum. Optional fields with the titles of "Comments:", "Formula:", "MW:", "Synonym:", and "CAS#:" can be between the "Name:" and "Num Peaks:" fields.

When a spectrum is saved to a text file from the **NIST/EPA/NIH Main** or **Replicates Library**, it will also contain fields with the names "NIST#:" and "DB#:". The field with the name "NIST#:" is on the same line as the "CAS#:" field and separated by a semicolon (;). "DB#" is same as the "ID#" displayed in the **Spectrum Text** Window. The mass/intensity list begins on the line following the line with the Field title "Num Peaks:". The lines of the mass/intensity list have no Field title.

When a file is created with all the allowable fields, each Field title should be on a separate line. Use the following format. The format for each spectrum in a text file usable by the **Program** should be as follows:

**Line 1:   NAME: Molecule** (Required, up to 511 characters)  
**Line 2:   COMMENT: Run 23, 8/8/88** (Optional, up to 1023 characters)  
**Line 3:   FORMULA: C6H6** (Optional, up to 23 characters)  
**Line 4:   MW: 78** (Optional)

If the spectrum submitted to **Library spectrum Search** contains a CAS registry number with its appropriate prefix (**CAS:** ), the display of the spectrum in the **Plot**, **Compare**, and **Structure** Windows will have the structure of the compound with that CAS registry number if present in the **NIST/EPA/NIH Main Library**.

**Line 5:   CAS: 71-43-2** (Optional)  
**Line 6:   SYNONYM: Chemical name synonym** (Optional, may be repeated)

The actual mass spectral data (number of peaks in the spectrum) may begin on the next line. It does not actually matter what line this is, as long as it precedes the line that starts the mass/intensity pair data. The VERY NEXT line and subsequent lines MUST contain the paired mass/intensity values.

**Line 7:   Num Peaks: 18** (This prefix and the exact number of mass/intensity pairs is Required.)  
**Line 8:   26 430; 27 340; 28 40; 37 480; 38 611; 39 1411; 49 300; 50 1792;**  
**Line 9:   51 2052; 52 1962; 63 340; 73 160; 74 480; 75 180;**

...

**Line n:       76 721; 77 1401; 78 9806; 79 651;**

The peaks need not be normalized, and the masses need not be ordered. The exact spacing and delimiters used for the mass/intensity pairs are unimportant. For example, the peaks above could equally be presented as:

(26,430),(27,340),(28,40),(37,480),(38,611),(39,1411),  
(49,300),(75,180),(50,1792),(51,2052),(52,1962),(63,340),  
(73,160),(74,480),(75,180),(76,721),(77,1401),(78,9806),(79,651)

or each mass intensity pair could be on an individual line. The following characters are accepted as delimiters (except "|"): |space|tab|,|;|:|(|)|[|]|}|

The file can be given any valid file name; however, it is best to use the extension "MSP". This is the recognized default extension in the **MS Search Program** when user-library spectra are to be imported.

Several examples of NIST text format mass spectral files are installed together with the **MS Search Program**, namely SAMPLIB.MSP, UNKNOWN.MSP, and USERDEMO.MSP.

---

## Import and Export of Retention Indices with a Spectrum

Retention index (RI) format was designed to provide compatibility with existing EI libraries and import of a mass spectrum into NIST MS Search.

In general, each RI value must have GC Column Type associated with it. Currently, there are 3 specific Column Types (1-3 in the table below) and 2 other (4-5) column types. RI of these two column type may be compared with any selected specific Column Type:

	Column Type	Accepted Tags (case insensitive)
1	Semi-standard non-polar	s, db5, semi, SemiStandardNonPolar, ssnp, SemiStdNP
2	Standard non-polar	n, db1, nonpolar, snp, standard, StandardNonPolar, StdNP
3	Standard polar	p, pol, polar, sp, StandardPolar, stdp, StdP, StdPolar
4	Any (n-alkane standard)	a, all, AllColumnTypes, AnyColumnType, std, Any
5	Unspecified	u, un, uns, unspecified, User

When importing or adding a spectrum with RI of Unspecified column type to a library, there is an option of changing RI column type to any of the three specific Column Types.

### ***RI Export with a Spectrum in a MSP or JCAMP file***

A compound may have up to 3 RI values of different specific Column Types or a single RI of any of other column type. Actually used by NIST MS Search RI values are exported into a single line in a MSP file in the following format with **Accepted Tags** of up to 3 different specified Column types:

**Retention\_index:** SemiStdNP=654/11/226 StdNP=654/7/286 StdPolar=957/17/131

or

**Retention\_index:** Any=1000

or

**Retention\_index:** User=2059

Slash-delimited numerical values following **Accepted Tags** are

*Experimental RI median/deviation/number of datapoints*

In NIST MS Search Spectrum Text Information window, RI values of the first example (Benzene) are displayed as

[Experimental RI median±deviation \(#data\)](#)

Semi-standard non-polar: 654±11 (226)

Standard non-polar: 654±7 (286)

Polar: 957±17 (131)

Other Unspecified RI values possibly present in the Comment (see the next section) are left unchanged in the exported spectrum.

Single RI value<sup>6</sup>, if available, is exported into a JCAMP file as

**##\$RETENTION INDEX=654**

---

<sup>6</sup> If RI values of multiple specified Column Types are associated with a spectrum, than RI of the first available column type in the following sequence is output into a JCAMP file:

- Semi-Standard Non-Polar
- Standard Non-Polar
- Standard Polar

---

## ***RI Import with a Spectrum in a MSP or JCAMP file***

Legacy MSP file formats allow saving several possibly different RI values of Unspecified column type, which may contradict each other. NIST MS Search picks up only one of them according to the priority rules described below.

If the last out of all lines starting with **Retention\_index:** or **RI:** or **Synon: \$:29** in a MSP file has format described in the previous section with possibly different **Accepted Tags**, then contents of this line is used as input.

Otherwise, there may be up to 4 untagged RI value locations/formats of Unspecified column type. They are described in the following table (*n* is the RI value):

<b>Priority</b>	<b>Format</b>	<b>Location</b>
Highest	<b>RI:</b> <i>n</i>	Last occurrence in comment
	<b>RI=</b> <i>n</i>	Last occurrence in comment
	<b>RI:</b> <i>n</i>	Last occurrence in comment
Lowest	<b>Retention_index:</b> <i>n</i> or <b>Retention_index=</b> <i>n</i> or <b>RI:</b> <i>n</i> or <b>RI=</b> <i>n</i> or <b>Synon: \$:29</b> <i>n</i>	Last occurrence in MSP file

RI value imported from a JCAMP file always has Unspecified column type.

## APPENDIX 1: Creating AUTOIMP Files Used in NIST MS Search

It may be necessary to create AUTOIMP.MSD as the primary locator file used for the importation of spectra from third-party mass spectral data systems and/or AUTOIMP.STR for interaction between the **MS Search Program** and third-party drawing programs. These files may be created using Notepad with certain versions of Microsoft Windows.

If Microsoft Notepad is used, the default extension is .TXT (Text Document) even though AUTOIMP.MSD or AUTOIMP.STR has been entered in the "File name:" field. The default settings for Windows Explorer will not show this .TXT extension. The file will not work unless the .TXT extension is removed. The file can be renamed by highlighting the file in Windows Explorer and selecting **Rename** from the **File** menu.

When saving the file using **Save As** option, this problem may be completely avoided by selecting **All Files** in the **Save as type** drop-down list box (Figure 44). In this case, the file will be saved with the extension provided in the "File name:" field (see the snapshot of the **Save As** dialog box below).

When saving the file using the Notepad **Save As** option, also make sure the **ANSI** Encoding is selected (in some versions of Windows, the Encoding option is not present).

After File name, Save as type, and Encoding are set correctly, press the **Save** button.

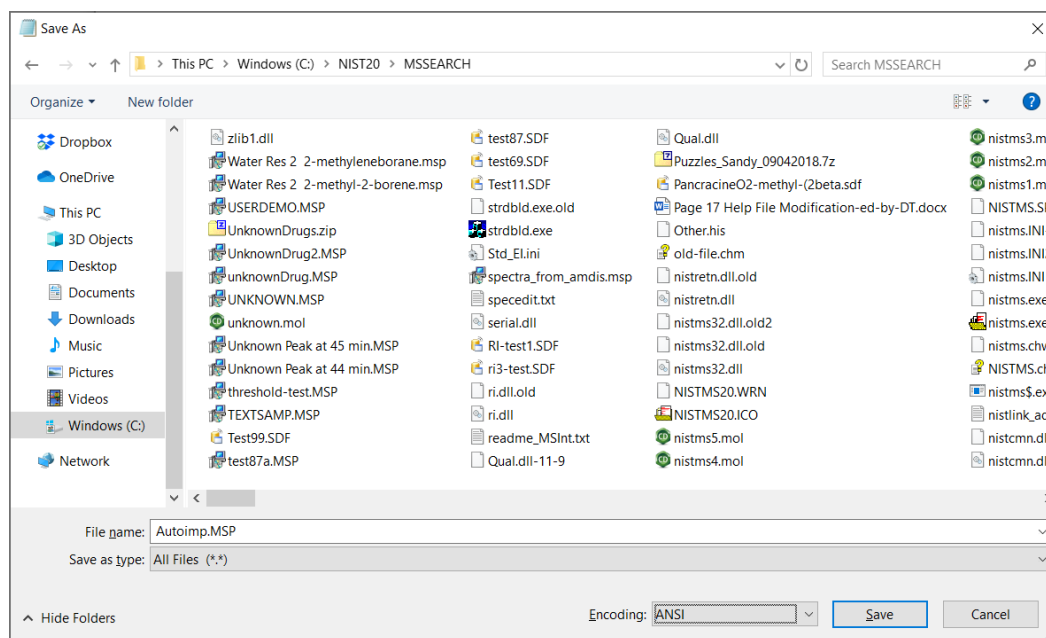


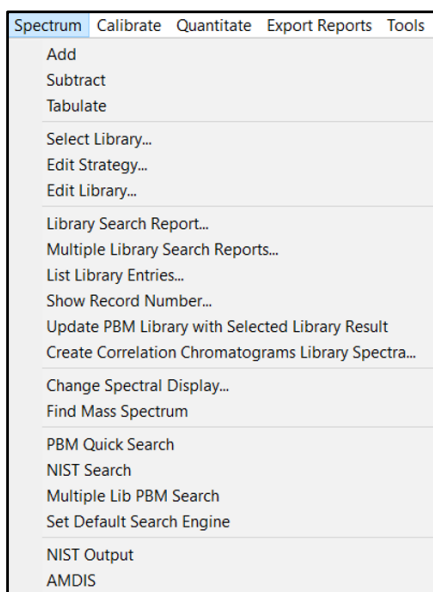
Figure 43. Save As dialog box displayed within Windows Notepad.

## APPENDIX 2: Using the NIST MS Search Program with ChemStation

The **NIST MS Search Program** has been supported through several versions of the HP/Agilent Technologies ChemStation. In the recent versions of the ChemStation for GC/MS, both the **NIST Output** and the **AMDIS** options will appear on the **Spectrum** menu of the Data Analysis software when using the Enhanced Quantitation, Standard Quantitation, or Gasoline mode; and on the **Lib.** menu when using the Environmental mode if the appropriate macros have been installed. The **NIST/EPA/NIH MS Library (NIST 20)** setup will install the two required macros along with additional macros that will allow for greater flexibility with the ChemStation software.

It is possible to use the ChemStation double-click option on a displayed spectrum to send that spectrum to the **NIST MS Search Program's Spec List Window**, bringing the **MS Search Program** into focus. If the appropriate options (including **Automation**, see Figure 25 and Figure 28) have been set in the **NIST MS Search Program** then, after the spectrum is automatically prepended to the **Spec List Window**, it will be searched.

If using the Enhanced Quantitation mode on any version of ChemStation, the setup will automatically add four new options to the **Spectrum** menu (**PBM Quick Search**, **NIST Search**, **Multiple Lib PBM Search**, and **Set Default Search Engine**). Each of these menu options is explained below. If these options are to be added to other modes of the ChemStation software, see the **MENU\_MOD** document file on the **NIST MS Search Program** installation CD ROM in the \support.cs directory. This document is also available after installing **NIST 20 MS Library** by selecting **ChemStation Menu Modification** from the **NIST Mass Spectral Database** program folder.



**Figure 44. Spectrum menu as displayed in ChemStation Enhanced data analysis.**

Some older versions of **ChemStation** will have **ASCII Output** as opposed to **NIST Output**, the **Spectrum** menu, or the **Lib.** menu found on the Environmental mode. Selecting either the **NIST Output** or **ASCII Output** results in the display of the **Options for NIST Library Search** dialog box with six *radio button* options and **OK** and **Cancel** buttons. The Macro NISTINT.MAC must be installed in the HPCHEMMSEXE or MSDCHEMMSEXE directory for these menu selections to be displayed.

When all the necessary macros have been installed in the **ChemStation's** MSEXE directory and the modifications have been made to the ChemStation macros, the following options will be available from the ChemStation Spectrum menu (Figure 45):

The **PBM Quick Search** (PBMSrch.mac) performs the normal ChemStation PBM search of the currently displayed spectrum.

The **NIST Search** (NISTSrch.mac) passes the currently displayed mass spectrum to the **NIST MS Search Program**. The spectrum will be prepended to the **NIST MS Search Program's Spec List Window**. If the **NIST MS**

**Search Program** is configured to search automatically, then the search will be started when the spectrum is passed to the **NIST MS Search Program**. Otherwise, double clicking on any spectrum in the **Spec List Window** located in the **Lib. Search** tab to start a search.

The **Multiple Lib PBM Search** (LS.mac) performs a **ChemStation PBM Search** using up to three different libraries (all of which have to be specified using the dialog box displayed by selecting the **Select Library** option on the **Search** menu). The top 20 Hits from each library will be displayed in order of decreasing match quality.

**Set Default Search Engine** (SWITCH.MAC) displays a dialog box that allows the selection of the default library search to be evoked when double clicking the RMB menu on the currently displayed spectrum. The **NIST MS Search** is the default until another search type is set as the default.

**NIST Output** (NISTINT.mac) displays a self-explanatory dialog box that allows for the use of the **NIST MS Search Program** in a number of different ways and the copying of ChemStation user libraries as **NIST MS Search Program** user libraries.

**AMDIS** (AMDIS.mac) loads AMDIS with the data file currently being viewed with the ChemStation Data Analysis directly into AMDIS for analysis. Once in AMDIS, select **Run** to perform the analysis.

The **Lib2NIST** converter program will be installed along with the **NIST MS Search Program**. This program allows mass spectral libraries that have been created using the internal utilities of the ChemStation software (including structures) and other mass spectral libraries to be copied to the NIST user libraries format that can be employed by the **NIST MS Search Program**. This utility is evoked as an option from the **Options for NIST Library Search** dialog box displayed by selecting the **Copy ChemStation Lib to NIST Format** *radio button* and then selecting the **OK** button or by selecting the **Lib2NIST** icon in the **NIST Mass Spectral Database** program folder.

An additional macro is found on the **NIST MS Search Program's** installation disk, **AUTONIST.MAC**. This macro is to be incorporated into automated sequences. It automatically selects spectra based on the autointegration of the current data file and switches to the **NIST MS Search Program** where the spectra are automatically imported into the **Spec List** Window. If the proper settings have been made in the **Program**, these spectra are searched and the focus returned to the HP MS software. These features can only be utilized with G1034C, Version C.3.00.00 or later of ChemStation.

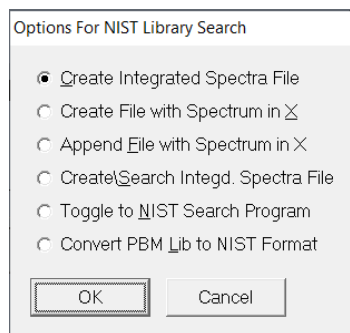
### Additional Information Regarding Modes Other Than Enhanced Quantitation

During the installation of the **NIST/EPA/NIH MS Library (NIST 20)**, if the **NIST MS Search Program** has been selected as the ChemStation's default search engine, each time a spectrum is double clicked, the displayed spectrum, will be prepended in the **NIST MS Search Program's Spec List** Window. The **NIST MS Search Program** will be brought into focus *regardless* of the data analysis mode (Enhanced Quantitation, Standard Quantitation, Gasoline, or Environmental). One of the other above-referenced search methods can be changed to by selecting the **ChemStation Default Search Mode** icon in the **NIST Mass Spectral Database** program folder.

MODIFICATION OF MACROS IS NOT NECESSARY TO SWITCH FROM THE NIST MS SEARCH PROGRAM AS THE DEFAULT SEARCH METHOD FOR THE DOUBLE-CLICK FEATURE OF THE ChemStation.

### HP\_NIST MACRO (NISTINT.MAC)

When **NIST Output** is selected from the **Spectrum** menu of the MS ChemStation or **ASCII Output** from the **Lib** menu of the Environmental software, the **Options For NIST Library Search** dialog box is displayed (Figure 46). This dialog box consists of six options and **OK** and **Cancel** buttons. Each option



**Figure 45. Result of selecting NIST Output from the Spectrum menu.**

has a *radio button* next to it. Use the Mouse pointer and the left Mouse button (LMB) on the *radio button* to select an option. Selecting the **OK** button results in execution of the option.

Items in the **Options For NIST Library Search** dialog box are selected by putting the Mouse pointer on the appropriate *radio button* and clicking the LMB. Then select the **OK** button to execute the desired option. If any of the first three options is executed, the dialog box will be redisplayed.

Each time one of the top four menu selections is executed, the file HPNIST.TXT is created in the ChemStation HPCHEM or MSCHEM directory. This file contains the name and location of the text file that contains the saved spectra in the NIST format. The HPNIST.TXT file is deleted every time the **NIST MS Search Program** is brought into focus and spectra are automatically imported into the top of the **NIST MS Search Program's Spec List** Window.

---

At the same time the HPNIST.TXT, the second locator file, is created, a text file named TEMPSPEC.MSP containing one or more spectra is created or appended. The text file is located in the ChemStation directory HPCHEM or MSDCHEM. This text file will remain there until it is overwritten or removed manually.

When the **NIST MS Search Program** comes into focus, the spectra that have been saved in the ChemStation software are automatically copied to the **Spec List** Window. If spectra are already in the **Spec List** Window, a prompt to **Prepend** or **Overwrite** the current contents will be displayed. If **Prepend** is selected, the new spectra are added to the top of the **Spec List** Window (prepended). If **Cancel** is selected, the automatic import will be aborted. However, the HPNIST.TXT file may be deleted even in the case of an abort. If the options of **Automatic Search On** and **Return Focus to Caller upon Completion** have been selected in the **Automation** dialog box, selecting **Cancel** returns the focus to the ChemStation software.

### Menu “Options for NIST Library Search”

The **Create Integrated Spectra File** option runs the ChemStation’s Autointegrate routine on the displayed chromatogram and creates the text file containing the spectra that represent the apex of each chromatographic peak.

The **Create File with Spectrum in X** option creates a text file that contains the spectrum that is in X of the ChemStation Stack display. This is where background-subtracted and averaged spectra will be found. This is the currently displayed spectrum.

The **Append File with Spectrum in X** option appends the spectrum in X of the Stack display to the existing text file. If a text file does not exist, one will be created.

The **Create/Search Integd. Spect. File** option creates a text file that contains spectra representing the tops of each chromatographic peak and either starts or switches to the **NIST MS Search Program** where the spectra are automatically imported into the **Spec List** Window. If **Automatic Search On** in the **NIST MS Search Program’s Automation** dialog box has been selected, each spectrum will be searched. The results are optionally printed, and the results of each search are stored in the **Hit List** History. If **Return Focus to Caller upon Completion** has been selected in the **Automation** dialog box, then after the last spectrum is searched and printed, the focus will be returned to the ChemStation software. If spectra are already in the **NIST MS Search Program’s Spec List** Window, the **Append/Overwrite/Cancel** query will be displayed.

The **Toggle to NIST Library Program** option either starts the **NIST MS Search Program** or switches to it if it is already running. The ChemStation software can be redisplayed by holding down the <Alt> key and pressing the <Tab> key, or the **NIST Program’s Switch To Caller** button can be selected. The **NIST MS Search Program** can be brought back into focus by using the <Alt><Tab> command or by selecting the **Toggle** option from the **Options For NIST Library Search** dialog box.

The **Copy ChemStation Lib to NIST Format** option is described in **APPENDIX 3**.

### AUTONIST Macro

This is a non-interactive macro that can be executed using the command line instruction:

MACRO “AUTONIST.MAC”, go

or from within the Method. Every Method directory contains a file DEUSER.MAC. Edit this file to contain the string:

MACRO “AUTONIST.MAC”,go.

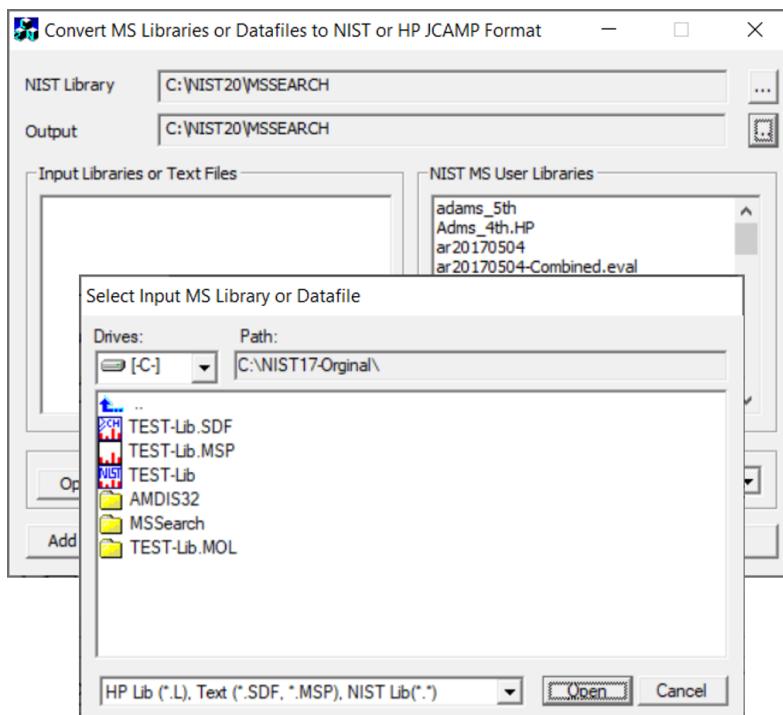
This File is in the \SUPPORT.HP directory of the CD ROM and must be copied to the \HPCHEMMSEX directory.

When the Method is run, the AUTONIST Macro will execute. The chromatogram is autointegrated, and the spectra representing the top of each chromatographic peak are stored in a text file. The focus is then switched to the **NIST MS Search Program**. The spectra are imported into the **Program’s Spec List** Window. If spectra are already in the **Spec List** Window, they will be overwritten.

If **Automatic Search On** and **Return Focus to Caller upon Completion** have been selected in the **NIST MS Search Program’s Automation** dialog box, the spectra will automatically be searched. After the last spectrum is searched, the focus is returned to the HP MS software. For this macro to function properly, it will be necessary to have selected the above two options in the **Automation** dialog box.

## APPENDIX 3: Copy ChemStation and Other Database to NIST User Library Format

This is a separate program called by selecting the **Lib2NIST Converter** icon on the **NIST Mass Spectral Database** program folder. This program allows mass spectral libraries in a variety of formats (along with structures or structures in separate files) to be selected and copied to the NIST user library format. For all 32-bit Windows (Windows NT, Windows 98/Me/2000/XP/Vista/Windows 7, 8, 8.1) and Windows 64-bit platforms, the **Lib2NIST** converter program replaces the **HP2NIST** program that was provided with previous versions of the **NIST MS Search Program**. The use of the **HP2NIST** program is described at the end of this section. Another feature of the **Lib2NIST** program is the ability to take user libraries that were developed by using versions 1.5 through 1.7 of the **NIST MS Search Program** and copy them into a format fully compatible with the current version of the **NIST MS Search Program**.



**Figure 46. Opening display of Lib2NIST program. When started, the Lib2NIST converter program displays a file dialog box overlaid on the program's screen.**

Select the library to be copied and then click on the **Open** button (Figure 47). The drop-down arrow will also allow for selecting from a variety of JCAMP formats. More than a single library can be selected in this dialog box by using the standard Windows multiple file selection techniques.



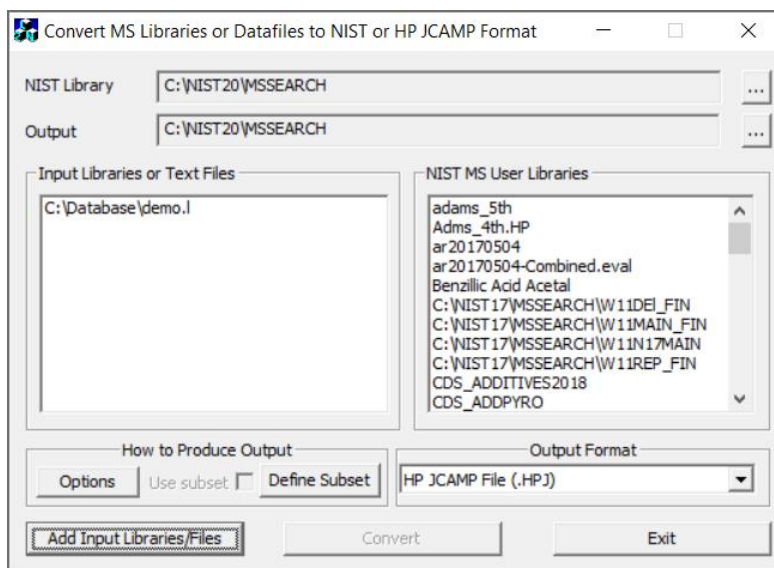


Figure 47. Lib2NIST converter program display.

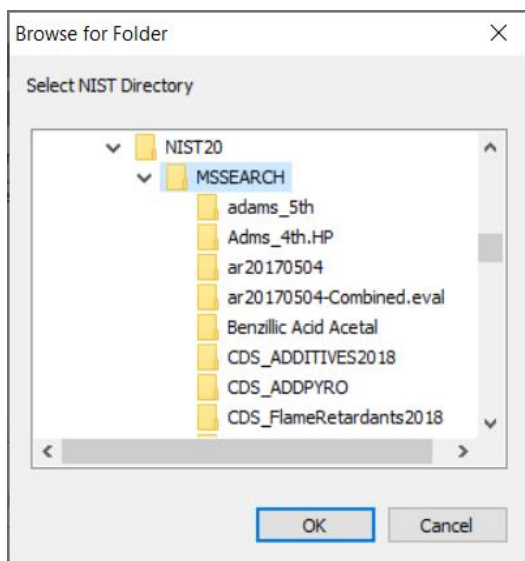


Figure 48. Conversion output selection dialog box.

The selected libraries will now be listed in the **Input Libraries** or **Text Files** window of the **Program's** display (Figure 48).

The **NIST Library** and **Output** locations can be changed by selecting the appropriate button located to the right of the two fields. This will result in the display of the **Browse for Folder** dialog box (Figure 49).

Select the desired directory (folder) and then click on the **OK** button to change the location.

The **Define Subset** button will display a dialog box that allows for the choice of a range of input spectra ID or CAS numbers. The **Use subset** check box is grayed, unless an entry has been made in the **Define Subset** dialog box (Figure 49).

After making any desired entries in the **Define Subset** dialog box, select the **OK** button to return to the **Program's** main display.

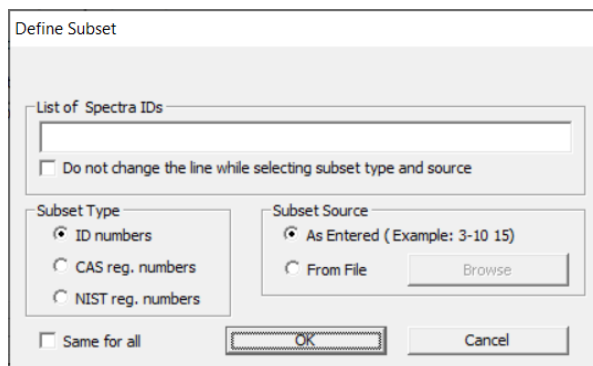


Figure 49. Define Subset dialog box.

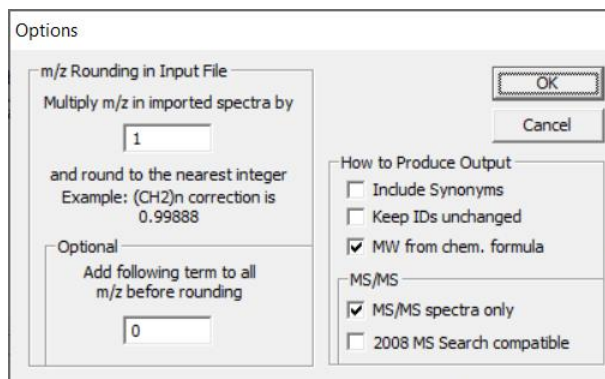


Figure 50. Options dialog box display.

The **Options** button on the **Program's** display will display the dialog box shown above in Figure 50. This allows for the entry of mass defect corrections to be applied to  $m/z$  values of imported spectra and whether to: 1) include synonyms from the source file, 2) retain the ID numbers (or sequence numbers) from the source library or file<sup>7</sup>, 3) calculate molecular weights (nominal mass) from formula, and 4) create or convert MS/MS and/or In-source libraries<sup>8</sup>.

Select the libraries to be converted in the **Input Libraries or Text Files** window of the **Program's** display, make sure the Output Format is selected correctly, and then click on the **Convert** button. The selected libraries will then be converted and placed in the specified output directory. Advanced features of Lib2NIST are explained in the file CMDLINE.pdf installed with the Lib2NIST. Transliteration rules of extended ASCII characters are listed in the ASCII text file HPTRANS.TBL and may be modified in this file by the user.

<sup>7</sup> "MS/MS Spectra only" option disables "Keep IDs unchanged". To override this, use command line option /KeepIDs=YES

<sup>8</sup> In most cases, the "2008 MS Search compatible" option should be turned OFF. This option is intended only for saving MS/MS spectra into a MSP text file in a format, which is compatible with old versions of MS Search. MS/MS or In-source HiRes libraries created with this option have obsolete format unsuitable for reliable searching of high resolution spectra.

## APPENDIX 4: Using the NIST MS Search Program with Thermo Fisher Corporation Xcalibur Software

The Thermo Fisher Xcalibur software uses the NIST MS Dynamic Library (dll) as a library search engine. A copy of the **NIST MS Search Program** and **AMDIS** may be provided with a copy of the Xcalibur software. The **NIST/EPA/NIH Mass Spectral Library** may be optionally provided by the Thermo Fisher Corporation. The NIST tandem library is installed on all Thermo MS/MS instruments.

To add the **NIST/EPA/NIH Mass Spectral Libraries** to the Xcalibur software, run the **NIST 20 MS Library** setup. If it locates the following **NIST MS Search Program** and the **AMDIS** directories:

C:\Program Files\NISTMS\MSSEARCH

C:\Program Files\NISTMS\AMDIS32

then let the setup install **NIST 20** in these directories. The **NIST 20 MS Library** will then be properly installed along with any necessary program updates.

More recent versions of the Xcalibur software detect and use **NIST MS Search** installed in another directory (for example, c:\nist20\mssearch) and the **NIST/EPA/NIH Mass Spectral Library** using means explained in the **USE WITH INSTRUMENT DATA SYSTEMS** section of this manual.

Exporting MS/MS spectra to **NIST MS Search** transfers both the mass spectrum and the precursor  $m/z$  values.

It should be noted that the Qual Browser of Xcalibur exports only integer  $m/z$  values NOT accurate  $m/z$  values. Upon release of NIST 20, the only Thermo Fisher program capable of exporting accurate mass data from Xcalibur acquired spectra is the program *FreeStyle*.

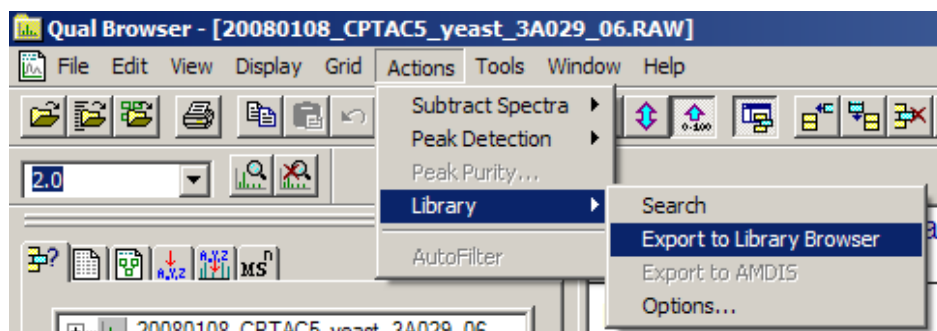


Figure 51. Exporting a spectrum to NIST MS Search.

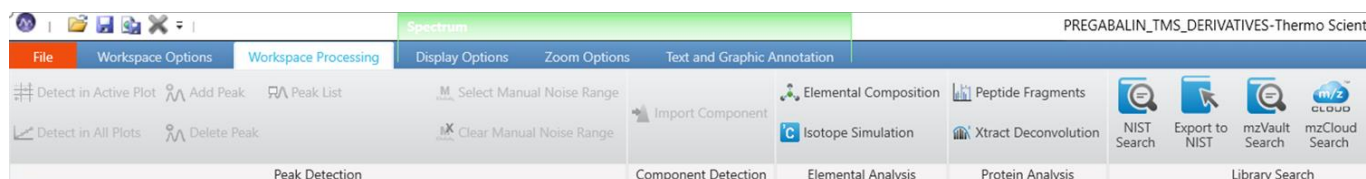


Figure 52A. Exporting a spectrum to NIST MS Search from Xcalibur's FreeStyle

---

## APPENDIX 5: Search Algorithms

There are two general ways that the **NIST MS Search Program** can retrieve library spectra that resemble the submitted spectrum. They are the **Identity Search** and the **Similarity Search**. An **Identity Search** is designed to find exact matches of the compound that produced the submitted spectrum and therefore presumes that the unknown compound is represented in the reference library. Only experimental variability prevents a perfect match. The **Similarity Search** is optimized to find similar compounds and is intended for use when a compound cannot be identified by the **Identity Search** (it is probably not in the library).

### Screening

For the sole purpose of achieving rapid retrieval rates before actually comparing spectra, modern search algorithms first identify a subset of library spectra with important features in common with the unknown spectrum. While this can vastly reduce search times, it can also screen out the correct spectrum. When this occurs, the correct retrieval cannot appear in the **Hit List**. Because of unavoidable blind spots in simple algorithms, even a closely matching spectrum can be excluded. This is like “throwing out the baby with the bathwater”. This is probably the most serious failure of any mass spectral search system.

A variety of filtering algorithms for Identity searching were tested in an effort to avoid throwing away correct matches without sacrificing performance. These algorithms were tested using the 12,592 spectra that comprised an older version of the **NIST/EPA/NIH Selected Replicates Library**. The best performing algorithm used “a ranked peaks in common” logic similar to that incorporated in existing data systems. This finds library spectra with the largest number of peaks in common with the unknown spectrum consistent with a required minimum number of identified spectra. Tests showed this minimum number to be about 50. This procedure retrieved 95% of the matching compounds (5% of matching compounds were lost). By scaling peaks by their  $m/z$  values, a 98% success rate was achieved at the same search speed. This is the screening logic of the **Quick Identity Search** option. By combining results using several screening criteria, a 99.4% success rate was achieved with a modest reduction in search speed. At this level, virtually all correct matches that were screened out were very dissimilar to the unknown spectrum and would have produced low Match Factors. Both peak scaling and merging of multiple screening results are used in the **Identity Search (EI Normal)**. A more detailed description follows:

**Peak Scaling:** The determination of the largest peaks in a spectrum was made after first multiplying the abundance of each peak by the square of its  $m/z$  value. The most intense peak in the scaled unknown spectrum is compared against the eight most intense peaks in scaled library spectra. The second most intense peak in the scaled unknown spectrum is then compared against the nine most intense peaks in the library spectra. This is repeated in decreasing order of intensity of peaks in the scaled spectrum of the unknown until the eighth most intense peaks are compared against the 16 most intense peaks in the scaled library spectra. This is the only screening procedure applied in **Identity Search (Quick)**.

**Merged Subsets:** Since the use of any single set of peak specifications was found to fail for certain classes of spectra, results of multiple sets of peak specifications were merged to reduce this problem. After extensive optimization studies, four separate peak specifications were selected. The first of these specifications is described above for the **Identity Search (Quick)**. The others were:

- 
- 1) The fourteen largest peaks in the scaled unknown spectrum were matched against the fourteen largest peaks in the scaled library spectra.
  - 2) The six largest peaks in the original (non-scaled) spectrum were matched against the six largest peaks in the original library spectra.
  - 3) The five largest peaks along with the “maximum mass” peak in both the unknown and library spectra were matched.

The screening algorithms for the **Similarity Search** are similar to the **Normal Identity Search** except that scaling and maximum mass peaks are not used. When neutral loss peaks are used in the **Hybrid Searches** and **Neutral Loss Search**, neutral loss peaks are used place of conventional peaks. For neutral loss peaks, abundances in library spectra are required to be within a factor of four of the abundances of corresponding unknown spectra (peak ranking and scaling are not used).

## Search

A mass spectrum can be represented as a row vector composed of the ordered peak intensities. It can also be considered to represent a single point in a multidimensional hyperspace defined by the  $m/z$  variables. Each of the intensities in the row vector represents the value of the coordinate of the spectral point along the individual mass axis in this hyperspace. If two spectra being compared are identical with respect to all the mass intensity pairs, their point representations in this hyperspace will coincide. If these spectra are very similar, their point representations will be close to one another. The Match Factor, which provides a sense of spectral similarity, may be regarded as the inverse of distance of the two point representations when each spectral vector has unit length.

The dot-product mass spectral search algorithm, which uses the cosine of the angle between the unknown and library spectral vectors, has been optimized by scaling peaks using the square root of their abundance. For the **Identity Search**, peaks were weighted by the square of their  $m/z$  value and a second term was added that compares ratios of adjacent peaks in library and unknown spectra. Its contribution was weighted so that it increased in importance as the proportion of common peaks increases. The only difference in the application of this algorithm to the **Identity Search** and **Similarity Search** is that  $m/z$  weighting is used only in the former. This defines the search algorithm used for both the **Similarity** and the **Identity Searches** in the **Program**.

For additional information, see Stein, S.E. “Optimization and Testing of Mass Spectral Library Search Algorithms for Compound Identification” *J. Am. Soc. Mass Spectrom.* **1994**, 5, 859–865.

---

## APPENDIX 6: Syntax for Some Constraints in Searches

### ***Syntax of Name Fragments Constraint***

Applies to: all searches  
Max. length: 39 characters  
First release: 2000

This "Name Constraint" field can contain up to eight name fragments.  
The name fragments are searched in compound main chemical name and in other names (synonyms).

The following examples explain the conditions when the name constraint is satisfied (spaces around && and !! are optional):

"COCA && INE": at least one name contains both "COCA" and "INE"

"COCA && !! INE": at least one name contains "COCA" and doesn't contain "INE"

"COCA !! INE": at least one name contains "COCA"; no name contains "INE"

### ***Syntax of Tags in Comment Constraint***

Applies to: all searches  
Max. length: 1023 characters  
First release: 2006

#### **General**

All information in NIST Peptide libraries is represented by the Tag=value convention. Typically, the "Peptide spectrum comment" field may contain the following items separated by spaces:

*Tag=string*

*Tag="string"*

*Tag=number*

Spaces may be inside a *string* only if the string is in double quotes. In older libraries, an unquoted *string* may also have space(s) inside a pair of parentheses ( ). Spaces are not permitted inside any *Tag* or next to the equal sign.

"Comment" fields in older libraries may also contain *Tag* not accompanied by the equal sign and *string*. All *Tag*, *string*, *substring*, and *word* comparisons described below **ignore letter case**.

---

## Entering Tags in Comment Constraint

Items in the “Tags in Comment” constraint field should be separated by spaces and/or linefeed characters. Unlike spectra comments, spaces next to the equal sign (as well as >, <, >=, <=, and :) are allowed inside the items.

Spectra returned must satisfy all entered constraints.

### Examples

	Item	Comments of the output spectra must have
1	"consensus"	A Tag or a space-delimited string <i>consensus</i>
2	<i>parent</i>	6 letter sequence <i>parent</i> anywhere in the Comment
3	<i>fullname="r.lecvk.c/2"</i>	Exactly this Tag=string:: <i>Fullname=R.LECVK.C/2</i>
4	<i>fullname=lec</i>	String following <i>Fullname=</i> contains sequence <i>LEC</i>
5	<i>Mods : ICAT_cl_hi</i>	String following <i>Mods=</i> contains word <sup>1)</sup> <i>ICAT_cl_hi</i>
The following rules apply to strings that are numbers		
6	<i>Diffmod=6.8</i>	String following <i>Diffmod=</i> begins with number 6.8, for example, <i>Diffmod=6.8/8</i> or <i>Diffmod=6.8</i>
7	<i>Diffmod=5.1:6.8</i>	String following <i>Diffmod=</i> begins with number between 5.1 and 6.8, for example, <i>Diffmod=6.8/8</i> or <i>Diffmod=6.8</i>
8	<i>Diffmod &gt; 6.8</i>	<i>Diffmod=</i> followed by a number greater than 6.8
	<i>Diffmod &gt;= 6.8</i>	<i>Diffmod=</i> followed by a number not less than 6.8
	<i>Diffmod &lt; 6.8</i>	<i>Diffmod=</i> followed by a number less than 6.8
	<i>Diffmod &lt;= 6.8</i>	<i>Diffmod=</i> followed by a number not greater than 6.8
<sup>1)</sup> <i>word</i> is a string delimited by any character but a letter, a digit, or an underscore “_”; If <i>word</i> contains spaces or equal sign it should be entered in quotes.		

To reverse the meaning of constraints 1–5 (that is, change to “spectra must not have”), enter caret ^ as the first letter:

1	^"consensus"	Tag <i>consensus</i> is not present
2	^ <i>parent</i>	Substring <i>parent</i> is not present
3	<i>fullname=^"r.lecvk.c/2"</i>	String tagged <i>fullname</i> is not exactly <i>r.lecvk.c/2</i>
4	<i>fullname=^lec</i>	String tagged <i>fullname</i> does not have substring <i>lec</i>
5	<i>Mods : ^ICAT_cl_hi</i>	String tagged <i>Mods</i> has not word <i>ICAT_cl_hi</i>

---

*More formal description*

	Item	Comments of the output spectra must have
1	" <i>Tag</i> "	Exactly this <i>Tag</i> is present in the comment, possibly not followed by the equal sign
2	<i>substring</i>	this <i>substring</i> anywhere in the comment
3a	<i>Tag</i> ="string "	exactly this <i>string</i> tagged with this <i>Tag</i> =
3b	<i>Tag</i> =""	This <i>Tag</i> has no <i>string</i> or its <i>string</i> is ""
4	<i>Tag</i> = <i>substring</i>	this <i>substring</i> in the <i>string</i> tagged with this <i>Tag</i> =
5a	<i>Tag</i> : <i>word</i>	exactly this <i>word</i> inside <i>string</i> tagged with this <i>Tag</i> =
5b	<i>Tag</i> : "word"	exactly this <i>word</i> inside <i>string</i> tagged with this <i>Tag</i> =
6	<i>Tag</i> = <i>number</i>	tagged string = <i>number</i> or <i>string</i> begins with <i>number</i>
7	<i>Tag</i> = <i>number</i> : <i>number</i>	tagged string <i>number</i> inside the given range
8a	<i>Tag</i> >= <i>number</i>	The number in the string less or equal to <i>number</i>
8b	<i>Tag</i> > <i>number</i>	The number in the string less than the <i>number</i>
8c	<i>Tag</i> <= <i>number</i>	The number in the string greater or equal to <i>number</i>
8d	<i>Tag</i> < <i>number</i>	The number in the string greater than the <i>number</i>

Notes.

- Inside an item, spaces are allowed next to signs =, >, <, >=, <=, or :. These signs as well as *Tag* and *string* or *number* or *word* may be located on separate lines.
- If a *Tag* created contains characters = or < or > or : then use double quotes to search for such a tag, for example, "<*Tag*>="string".
- *substring* in rows 2 and 4 may not contain characters < > = :.
- To search for a substring that may be interpreted as a number, for example, 1.23e5, use *Tag*==1.12e5 or *Tag*=="1.12e5" instead of *Tag*=1.12e5
- To reverse the meaning of constraints 1–2 (that is, change to “spectra must not have”), insert caret ^ in front of the "*Tag*" or *substring*
- To reverse the meaning of constraints 3–8 (that is, change to “spectra must not have”), insert caret ^ in front of the right hand side expression.
- To use caret ^ as the first character of a *substring* enter *Tag* == "^*substring*".
- In 3-8, *Tag* may be replaced with *Tag*?. In this case the spectrum is accepted not only if the condition(s), described in the above table are satisfied but also if *Tag* is missing.

Example of a multi-item constraint: The following items

Replicate - Line 2, substring  
 RT >= 11.2 - Line 8a, number is greater or equal  
 Sample = "yeast\_Nature\_CAM" - Line 3a, *Tag*="string"  
 Datfile = 011599.dat - Line 4, *Tag*=substring

fit the following “Comment” field:

Replicate RT=11.220m Sample="yeast\_Nature\_CAM" Datfile="F011599.dat"





---

*More formal description of reversed constraints*

	Item	Comments of the output spectra must have
1a	<code>^"Tag"</code>	Exactly this <i>Tag</i> is not present in the comment
1b	<code>~"Tag"</code>	Some of the tags are not exactly this <i>Tag</i>
2a	<code>^substring</code>	this <i>substring</i> is not anywhere in the comment
2b	<code>~substring</code>	this <i>substring</i> is not anywhere in the comment
3a	<code>^Tag="string"</code>	all tags but this <i>Tag</i> have exactly this <i>string</i>
3b	<code>~Tag="string"</code>	some tags but not this <i>Tag</i> have exactly this <i>string</i>
3c	<code>^Tag=^"string"</code>	all tags but this <i>Tag</i> haven't exactly this <i>string</i>
3d	<code>~Tag=^"string"</code>	some tags but not this <i>Tag</i> haven't exactly this <i>string</i>
3e	<code>Tag=^"string"</code>	this <i>Tag</i> has no string or its string differs from <i>string</i>
3f	<code>Tag=^""</code>	this <i>Tag</i> has a non-empty string
3g	<code>all=^"string"</code>	Each of the tagged strings has not exactly this <i>string</i>
3h	<code>Some="string"</code>	some tagged strings have exactly this <i>string</i>

Items 4–8 from the previous section (1.2.2) may be given reversed meaning in the same way as 3a–3f.

Reserved expression `all=` is same as `^""=` and `^=`

Reserved expression `some=` is same as `~""=` and `~=`

To address a tag named `all` or `some` put it in quotes, for example:

`"all"=string` `"some"=string`

Quoting tags allows searching for exact combination of tags, for example

`"Single Tryptic_simple Parent"=358.07` or `"Charge=2 Scan"=82`

`Tag?=^whatever` is considered found if either `Tag=^whatever` has been found or `Tag` is not present.

## Tags in Comment Constraint for MS/MS Text Information

Text information displayed for **NIST 20 MS/MS Library** mass spectra has special tagged text fields, for example:

```
MSn pathway: [M+H-H2O]+=>311.4
Precursor type: [311.4]+
Spectrum type: MS3
Precursor m/z: 375.2894, 311.4
Instrument type: IT/ion trap
Instrument: Thermo Finnigan Elite Orbitrap
Sample inlet: direct flow injection
Ionization: ESI
Collision gas: He
Collision energy: 35%
Ion mode: P
Notes: Consensus spectrum; Acetonitrile/Water/Formic acid; Vial_ID=1741
InChIKey: KXGVEGMKQFWNSR-UHFFFAOYSA-N
```

In MS Search v.2.2, **Tags in Comment** Constraint allows searching these tagged fields. Special tags are to be used for this purpose.

To enter such a special tag, type dollar sign (\$) immediately followed by the tag displayed in the spectrum text information, omitting spaces. Slashes, dashes, underscores in the tag may be omitted or added. Tags differing only in CAS, CAS#, or CASNO at the end of the tag are equivalent. The special tag and search string are case-insensitive.

The syntax is almost the same as the one described above for **Tags in Comment** Constraint.

Below are several examples.

To search for	Enter into the Tags in Comment Constraint
<i>precursor m/z</i> in range* 374.9–375.4	\$Precursor_m/z=374.9:375.4
<i>Instrument type</i> containing word IT	\$Instrument_type:IT
the exact <i>Instrument</i> name	\$Instrument="Micromass quattro"
spectra that have any Related CAS# entry	\$Related_CAS
spectra that have no Collision energy entry	^\$CollisionEnergy

\* Note: MS<sup>3</sup> spectrum with text information "Precursor m/z: 375.2894, 311.4" also will be found. That is, for MS<sup>n</sup> (n>2) spectra, the first precursor m/z (corresponding to MS<sub>2</sub>) will be found.

### List of all acceptable special tags

\$aux	\$ion_mode	\$precursor_m/z
\$charge	\$ion_mw	\$precursor_type
\$collision_energy	\$ion_name	\$pressure
\$collision_gas	\$ionization	\$related_cas#
\$compound_type	\$known_impurity	\$retention_index
\$cone_voltage	\$mass_range	\$ri
\$inchikey	\$maximum_intensity	\$salt
\$in-source_voltage	\$msn_pathway	\$salt/mix_cas#
\$instrument	\$notes	\$sample_inlet
\$instrument_type	\$peptide_mods	\$special_fragmentation

---

\$ion_formula	\$peptide_sequence	\$spectrum_type
---------------	--------------------	-----------------

Special tags inside parentheses are considered identical:

(\$aux \$notes) (\$in-source\_voltage \$cone\_voltage) (\$retention\_index \$ri)

Some of these tags may be found in other **NIST 20 EI Libraries**, for example, \$related\_cas#, \$inchikey.

See also MS Search Help topic "Spectrum Text Information"

## MS/MS Instrument Type Constraint

This feature will only work with **NIST 20 MS/MS Libraries** or similarly formatted user libraries.

It affects only ms/ms spectra (that is, spectra which have precursor  $m/z$  value); therefore, it ignores all EI spectra, which means applying this constraint to a search of the EI Library does not affect the Hit List.

This constraint has 7 options (*check boxes*) that may be combined to produce a Hit List containing such spectra that each spectrum fits one of the selected Instrument Types.

The "Instrument type" field in a MS/MS mass spectrum text information is interpreted in the following way:

Instrument type <i>check box</i>	"Instrument type" field in MS/MS mass spectrum
Ion Trap	"IT/ion trap" or "QQIT" or "QqLIT"
Q-TOF	"Q-TOF"
Ion Trap + FT	"IT-FT/ion trap with FTMS"
HCD	"HCD"
Triple quadrupole	"QqQ"
Other	any other than the above five instrument types
Unspecified	No instrument type specified

Selecting all seven would select all mass spectra (the same as if the constraint is disabled).

The version of **Lib2NIST** included in **NIST 20** or a later version is needed to produce such a library because the instrument type is saved in the **NIST 20 Tandem** mass spectrum record as a new binary field; therefore, this constraint would not work with the **NIST 08 Tandem Library** or even the **NIST Peptide Libraries** created with the previous version of **Lib2NIST**.

In case of NIST 11 or NIST 12 MS/MS libraries, Q-TOF check box would work as Q-TOF and HCD check boxes selected together, HCD and Ion Trap + FT would not find any spectra although 737 HCD spectra are present in each of NIST 11 and NIST 12 MS/MS library. However, HCD spectra in these libraries may be found by entering Tags in Comment constraint \$Instrument\_type=HCD or Name Fragment constraint \$:06HCD

See **Read Me Lib2NIST** (Lib2NIST.TXT or Readme\_Lib2NIST.txt) for more details on rebuilding MS/MS library.

## APPENDIX 7: Contributing EI Mass Spectra to the NIST/EPA/NIH Mass Spectral Library

Users of the **NIST MS Search Program** are invited to submit mass spectra for inclusion in the **NIST/EPA/NIH Mass Spectral Library**. Descriptions of both ideal and minimum acceptable data are provided below. Please provide data that is as close to the ideal as convenient.

Characteristic	Ideal	Minimum Acceptable
Substance	Pure compounds (single GC component) of confidently known structure that are commercially important, occur in the environment, or are a natural product (including analytically useful derivatives)	Any single chemical compound whose identity is confidently known
Method	Acquired by GC/MS at an ionization energy of 70 eV under conditions such that isotope ratios are not distorted and ions above the molecular ion are absent and processed using AMDIS or other deconvolution software	Acquired at an electron ionization of not less than 20 eV under conditions such that isotope ratios are not grossly distorted and significant ions above the molecular ion are absent; significant background subtracted
Requested Information:		
Contributor	Contributor's name and/or institution (as it is to appear with the spectrum)	Contributor's name and/or institution (as it is to appear with the spectrum)
Chemical name	Chemical name (preferably in Chemical Abstracts or IUPAC format)	Chemical name (or structure)
Chemical Abstracts Registry Number	Chemical Abstracts Registry Number	-
Chemical structure	Chemical structure (in MOL/SDF file format)	Chemical structure (or name)
Synonyms	Alternative names (trade names, trivial names, etc.)	-
Mass spectrum	Mass spectrum (unit $m/z$ resolution or better) with all observed nominal masses from no higher than $m/z$ 12 to above the molecular ion with intensities of 0.1 % or greater and intensity in electronic format (preferably as an NIST/EPA/NIH user library). Ions at half-mass should be kept in the spectrum and rounded down to the nearest mass.	Mass spectrum (unit mass resolution) with all observed nominal masses beginning no higher than one-third the molecular mass to above the molecular ion with intensities of 2.0%. Ions at half-mass should be kept in the spectrum and rounded down to the nearest mass.
Instrument type	Instrument type (magnetic sector, linear quadrupole, TOF, etc.)	Instrument type (magnetic sector, linear quadrupole, TOF, etc.)
Instrument model	Instrument manufacturer and model name	-
Ion source temperature	Ion source temperature (°C)	-
Inlet	Inlet (GC, probe, etc.) and inlet temperature (of probe, reservoir, etc.)	-

Retention index	Kováts Retention Index (with the commercial name of the GC phase or, if not available, whether the phase is polar or non-polar)	-
Impurities	Note masses not belonging to the compound and their origin, if known.	Note masses not belonging to the compound, if known.

---

## APPENDIX 8: Support Contacts

In our continuing commitment to quality, the NIST Mass Spectrometry Data Center is always looking to improve the quality of our Mass Spectral Libraries and programs for accessing them. Comments or questions about the quality of either or other Standard Reference Databases available from the NIST Standard Reference Data Program should be made by contacting:

Mass Spectrometry Data Center  
Biomolecular Measurement Division 645.04  
National Institute of Standards and Technology  
100 Bureau Drive, Mail Stop 8362  
Gaithersburg, MD 20899  
email: [massspec@nist.gov](mailto:massspec@nist.gov)  
website: <http://chemdata.nist.gov>

**NIST provides updates and enhancements to the NIST Mass Spectral Search Program, MS Interpreter and AMDIS. They can be downloaded from <http://chemdata.nist.gov>.**

