

# NIST Standard Reference Database 1A

---

## NIST/EPA/NIH Mass Spectral Database (NIST 11) and NIST Mass Spectral Search Program (Version 2.0g)

For Use with Microsoft® Windows  
User's Guide

---

### The NIST Mass Spectrometry Data Center

Stephen E. Stein, Director

#### *Evaluators:*

Anzor Mikaia, Principal  
Edward White V (EI)  
Vladimir Zaikin (EI)  
Damo Zhu (EI)  
O. David Sparkman (EI)  
Pedatsur Neta (MS/MS)  
Valeri Babushok (RI)  
Igor Zenkevich (RI)

#### *Programmers:*

Peter Linstrom  
Yuri Mirokhin  
Dmitrii Tchekhovskoi  
Xiaoyu Yang

#### *Documentation and Help:*

W. Gary Mallard  
O. David Sparkman  
Joan A. Sparkman

May 2011

### 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.

---

© 1987, 1988, 1990, 1992, 1994, 1995, 1996, 1997, 1998, 1999, 2000, 2001, 2002, 2005, 2008, 2011 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., XVT Software, Inc., 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

Installation .....	5
OS Compatibilities .....	6
Introduction.....	7
New Features in NIST 11 and the MS Search Program.....	9
NIST 11.....	9
GC Methods / Retention Indices Database.....	9
The NIST MS/MS Database .....	9
Version 2.0g of the MS Search Program .....	11
Previous version, 2.0f, of the MS Search Program .....	12
Tab Displays.....	13
Title Bar, Tool Bar, and Status Bar .....	18
Neutral Loss Display .....	20
The Lib. Search Tab .....	21
The Other Search Tab .....	21
The Name Tab.....	23
The Compare Tab.....	24
The Librarian Tab .....	24
Spectral Analysis Utilities for NIST 11 .....	25
Automated Mass Spectrometry Deconvolution and Identification System (AMDIS) .....	25
MS Interpreter.....	26
Example Searches .....	29
User Spectrum and Structure Searches .....	29
Automations .....	32
Limits.....	32
Constraints .....	32
The Search Menu (displayed by Selecting Search on the Menu Bar) .....	33
Search by Any Peaks.....	33
Exact Mass Search.....	36
Search by Name .....	37
Sequential Method.....	37
Basis for Interpretation of the Database Search Results .....	38
Substructure Information .....	39
Example of the Use of Substructure Information in Making Determination about an Unknown Mass Spectrum obtained by using CAD.....	41
Use with Instrument Data Systems .....	42
Use with Third-Party Drawing Programs .....	44
User Databases.....	45
NIST Text Format of Individual Spectra.....	47
APPENDIX 1: Creating AUTOIMP Files Used in NIST MS Search V.2.0.....	48
APPENDIX 2: Using the NIST MS Search Program with ChemStation.....	49

---

APPENDIX 3: Copy ChemStation and Other Database to NIST User Database Format .....	52
APPENDIX 4: Using the NIST MS Search Program with Thermo Fisher Corporation Xcalibur Software .....	55
APPENDIX 5: Search Algorithms .....	56
Screening.....	56
Search .....	57
APPENDIX 6: Syntax for Some Constraints in Searches.....	58
Syntax of Name Fragments Constraint.....	58
Syntax of Tags in Comment Constraint .....	58
Tags in Comment Constraint for MS/MS Text Information.....	62
MS/MS Instrument Type Constraint.....	63
APPENDIX 7: Contributing EI Mass Spectra to the NIST/EPA/NIH Mass Spectral Database.....	64
APPENDIX 8: Support Contacts.....	65

# Installation

## Windows NT 4.0, Windows 2000, Windows XP, Windows Vista, Windows 7

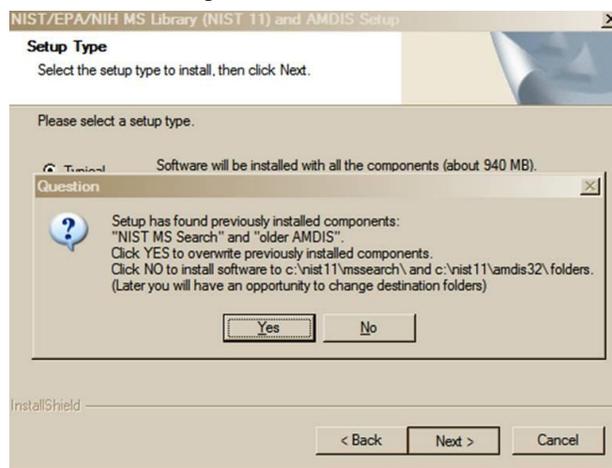
- Select the **Start** button on the lower left of the screen. Then select **Run** from the **Start** menu.
- Assuming the CD-ROM drive is D, type D:\SETUP.EXE and select **OK**. If the CD-ROM drive is a device other than D, replace D with the appropriate drive letter.
- Follow the instructions presented by the SETUP program.

In upgrading from a previous version of the NIST MS Database, there is a possibility to keep both versions.<sup>1</sup> The Installation Program may behave differently under Windows 7 and Vista due to User Account Control security features.

- Note: **AMDIS** is not compatible with Windows NT 4.0.

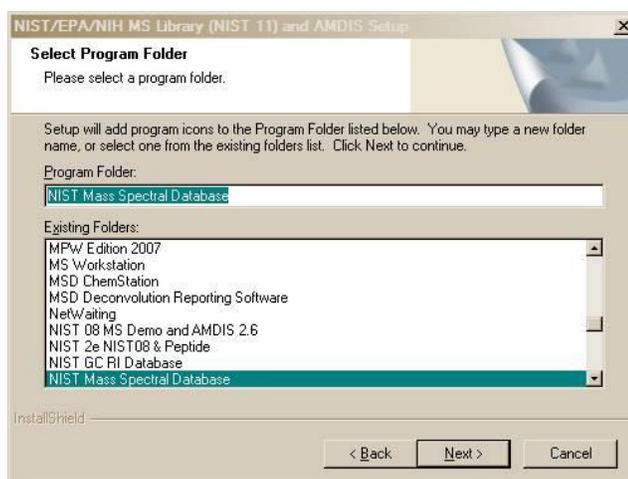
<sup>1</sup> In upgrading from a previous version, the Database and Programs will be overwritten by default. To avoid this, (1) the Database should be installed in a new folder, and (2) a new Program Folder should be selected for the Start Programs shortcuts.

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



**Question dialog box: Select overwriting the previous installation, or retain the previous installation and install NIST 11 in a new folder.**

(2) When the dialog box shown below appears, the name of the Program Folder must be changed to **NIST 11 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 11** and **MS Search** are **NOT** compatible with Windows 3.x, 95, 98, Me, NT 3.x

**NIST 11** and **MS Search** are fully compatible with Windows 2000, XP, Vista, and Windows 7 (XP, Vista, and Windows 7 in both 32- and 64-bit editions)

**NIST 11** and **MS Search** are partially compatible with Windows NT 4.0. Only **NIST MS Search**, which provides access to **NIST 11 MS Database**, will work. **NIST 11** version of **AMDIS** is not compatible with Windows NT 4.0. Also needed are:

- **Adobe PDF** reader to access the **NIST 11** 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 11** and **MS Search** have been tested with.

The latest version of **AMDIS** known to be compatible with Windows NT 4.0 SP1 and **MS Search** as distributed with **NIST 11** is version 2.64 (build Jan. 13, 2006).

---

## Introduction

The National Institute of Standards and Technology (NIST) introduced a version of its search software used with the NIST / U.S. Environmental Protection Agency (EPA) / National Institutes of Health (NIH) Mass Spectral Database for Microsoft Windows in 1995. At that time, the **NIST Mass Spectral Search Program for Windows** (the **Program**) would allow for various types of comparisons of an acquired or unknown spectrum with the **NIST/EPA/NIH Database**, its Database of Replicate Spectra, and user databases. The **NIST/EPA/NIH Database** was also searchable by multiple names that were associated with the spectra. This **Incremental Name Search**, which first appeared in a DOS version of the **Program**, was what analysts had been looking for since commercial mass spectral databases became available with mass spectrometry data systems: "how to retrieve a spectrum of a compound without having to search an acquired spectrum against the Database." Another unique feature of the **Program's** spectral comparison algorithms, even in the DOS version, was that the spectra selected from the database during the Presearch had all of their peaks compared with all of the peaks in the sample spectrum. All other commercial mass spectral database search systems at that time still used condensed spectra.

Other features retained in v.2.0 of the **Program** that were in the original DOS and Windows versions are the ability to retrieve spectra using **Molecular weight** (nominal mass), **Formula** (elemental composition), **Chemical Abstract Services Registry numbers (CASrn)**, and **Database ID number**. All these search methods, along with the versatile **Any Peaks Search**, are available for use with the **NIST/EPA/NIH MS Database** and user databases. In addition, multiple constraints can be added to searches that would limit the retrieved spectra population. The first Windows version introduced the ability to search multiple databases simultaneously.

Since the introduction of the first Windows version of the **Program**, other features have been added: 1) mass defect correction of  $m/z$  values in imported spectra, 2) ability to associate structures as MOL files with user database spectra, 3) ability to have name synonyms associated with spectra in user databases, 4) ability to use the **Incremental Name Search** with user databases, 5) ability to evaluate a **Hit List** from a **Neutral Loss** or **Hybrid Neutral Loss Similarity Search** to determine the probability of the presence and absence of various substructures, 6) ability to estimate the nominal mass of an unknown along with the ability to determine the number of atoms of chlorine and/or bromine that may be present in an analyte that produced a mass spectrum, 7) addition of the mass spectral interpretation tools **Isoform** and **MS Interpreter**, 8) an **MS/MS Search**, 9) the ability to estimate retention indices for most of the compounds in the **Mass Spectral Database** and user libraries and, most recently, 10) the ability to sort **Hit Lists** (alphabetically, by number of synonyms, and by number of other databases containing the compound) and the **Spec List** (alphabetically), 11) perform an **Exact Mass Search**, 12) use **Exact Mass** as a constraint for spectral and other searches, 13) use "Exact Mass peaks" in the **Any Peaks Search**, and 14) the ability to view and search the spectra and other information in the MS/MS Database that is a part of NIST 11. In addition, NIST 11 is provided with a GC Methods/Retention Index Database of nearly 350K records for over 70K compounds and an MS/MS Database of over 95K spectra. Since the introduction of the first Windows version of the **Program**, five new versions of the **NIST/EPA/NIH EI Mass Spectral Database** (NIST 98, NIST 02, NIST 05, NIST 08, and NIST 11) have been introduced resulting in an increase from ~129K spectra of ~108K compounds to ~243K spectra of ~212K compounds; almost double in spectra and compounds. Through the tireless efforts of the NIST Mass Spectral Data Center, the Database has undergone continued improvements in spectral quality through the replacement or deletion of spectra found not to meet the Center's rigid requirements for spectral quality.

**Nomenclature Note:** Throughout the years, two terms have been used in conjunction with data collections of mass spectra, chemical structures, and related information compiled by NIST: **Database** and **Library**. In this document, the preferred term is **Database**. A Mass Spectral Library is a collection of mass spectra and accompanying data (names, chemical structures, GC Methods/Retention Indices, etc.) which would be found in a printed volume. A Mass Spectral Database includes not only a collection of the spectra and accompanying data but also facilities for efficient searching, retrieval, and managing the information. The facilities include optimized data formats, indexing, and software. This describes the **NIST 11 Database** and the **Search Program** used with this Database.<sup>2</sup> Therefore, the term **Database** is more appropriate. Please be advised that the terms **Database** and **Library** are synonymous in the Manuals, Programs, and Help Screens. These terms may refer to the whole product (**NIST 11 Database**) or to its components (mainlib, replib, nist\_ri, nist\_msms, and nist\_msms2.)

---

<sup>2</sup> To be precise, this describes a database, a database management system with a low-level query interface, and a database application program.



---

# New Features in NIST 11 and the MS Search Program

## NIST 11

The 2011 version of the **NIST/EPA/NIH Mass Spectral Database (NIST 11)** contains four mass spectral databases and a Database of retention index (RI) values. There are two EI Databases, **mainlib** and **replib**.<sup>3</sup> Together they contain 243,893 spectra of 212,961 different chemical compounds. This is an increase of 20,853 compounds (~11%). The two MS/MS databases, **nist\_msms** and **nist\_msms2**, contain 95,409 spectra of 12,568 ions, a 6.4 and 2.3 times increase, respectively. The features of this database are described in more detail below. The GC Method/Retention Index Database contains 346,757 GC Methods with RI values for 70,835 compounds, a 61% increase in the number of compounds in NIST 08. The number of compounds with GC Methods/RI Database in the EI MS Database (**mainlib**) is 38,648, an increase of 76%.

Contents of the NIST/EPA/NIH (NIST 11) Mass Spectral Database		
Database Name	Database Description	Database Contents
mainlib	Main EI MS Database	212,961 spectra
replib	Replicate spectra Database	30,932 spectra
nist_msms	MS/MS Database	64,511 MS/MS spectra for 10,205 ions
nist_msms2	MS/MS Database	30,898 MS/MS spectra for 2,363 ions
nist_ri	Retention Index Database	346,757 Kovats RI values for 70,835 compounds

## GC Methods / Retention Indices Database

One of the items in the **Text Info** tab of the **Properties** dialog box (displayed by selecting **Properties** from the Right Mouse button (RMB) menu when the Mouse pointer in a **Text Info** Window is **Retention Indices**). Having a check in the box next to **Retention Index** will cause the experimental data for the retention index to be displayed with the spectrum. 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 two values. Displaying all values will cause the display to be slow in cases where there are a very large number of literature values. There are a total of 346,757 annotated GC Methods/Retention Index values for 70,835 compounds, which include literature source and measurement conditions. 38,648 compounds in the GC Method/RI Database are in the **mainlib** EI Database. In addition, for many compounds in the Database, an **Estimated Kovats Retention Index** will be displayed based on the structure associated with the spectrum. These values are reported in 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. Provided the text pane view has been set to display **Retention Indices**, these data will be displayed with the results of any search. If spectra in a user database have associated structures, an Estimated Kovats RI will also be listed for these spectra.

## The NIST MS/MS Database

MS/MS spectra, along with an increased number of fields relevant to MS/MS data, are provided in a set of files separate from the Main and Replicates data files that comprise the Database of EI spectra. The MS/MS spectra have been provided by contributors, measured at NIST, and extracted from the literature. The preparation of the

---

<sup>3</sup> Contents of NIST 08 library **nist\_salts** have been merged into NIST 11 databases **mainlib** and **replib**.

MS/MS Database revealed that at sufficiently high signal-to-noise measurement conditions, modern instruments are capable of providing very reproducible “database-searchable” spectra. Although collision energy is an important variable in Q-tof and related 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 MS/MS spectra in NIST 11 are provided in two databases, *nist\_msms* and *nist\_msms2*. The *nist\_msms* database is described in detail below. The second MS/MS database, *nist\_msms2*, contains spectra of compounds that do not have chemical structures displayed. They mostly are commercially available peptides. Spectra from the *nist\_msms2* database are not displayed in the **MSMS** tab. Otherwise, the format of the two databases is similar.

The **MSMS** tab allows access to the spectra in the *nist\_msms* database. The first time this tab is selected, the three panes of the display are blank with the exception of a single entry in the **ion list** pane, which is on the left of the tab’s display. This entry is the **MS/MS Database**. There is a box on the left that has a plus sign . Placing the Mouse pointer on this box and clicking expands to the first two levels (Positive Ions and Negative Ions). There are *plus boxes*  next to both the Positive Ion and the Negative Ion. Clicking on one of these *plus boxes* results in the display of the next level. This is a list of molecular weights of analytes for which a spectrum was obtained by MS/MS. Clicking on the *plus box* next to one of these molecular weight listings results in the display of a list of elemental compositions that will also have *plus boxes* next to them. Clicking on one of the elemental composition *plus boxes* will result in a display of a list of the compound names that have that elemental composition. In most cases, this is only a single compound. Clicking on the *plus box* next to the compound name will result in the display of a list of the spectra acquired under MS/MS conditions for this compound. If more than one spectrum was acquired, there is a separate entry for each spectrum. Clicking on the spectrum entry will result in displaying the spectrum information.

The listing of each spectrum begins with **ms1** or **ms2**. An **ms1** spectrum is composed of ions arising directly from the ion source region. They can arise from electrospray ionization (ESI) or atmospheric pressure chemical ionization (APCI). These ions are not subjected to collision-induced dissociation after exiting the ion source.

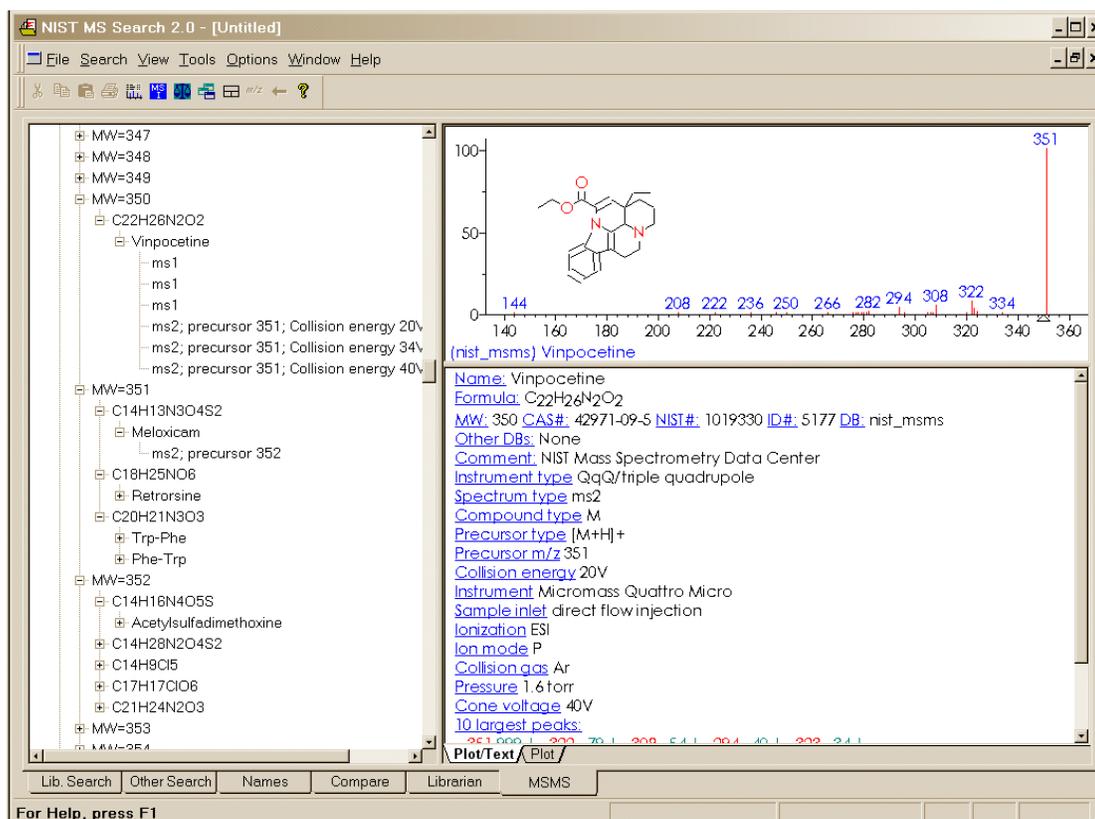


Figure 1. MSMS tab for data acquired using “triple quad”.

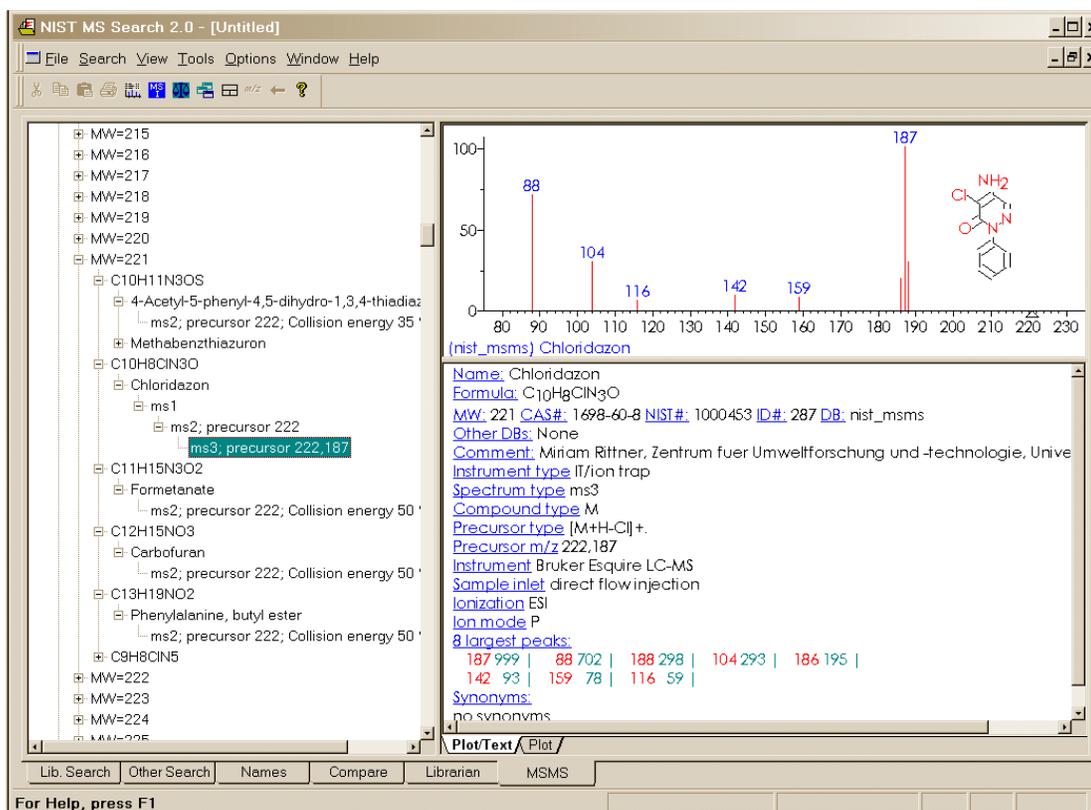


Figure 2. MSMS tab for data acquired using “ion trap MS<sup>3</sup>”.

For beam-type instruments (e.g., triple quadrupoles and Q-tof instruments), the **ms2** designation 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, the **ms1** may have a *plus box* next to it. Clicking on this *plus box* results in a subsequent spectrum listed as **ms2**, which may also have a *plus box*. Selection of the *plus box* next to the **ms2** spectrum can result in an **ms3** spectrum, which is obtained by the isolation of a product ion as a precursor ion for a subsequent iteration of MS/MS.

Clicking on the **ms3** results in the display of the spectrum as seen in Figure 2. The **ms2** designation is followed with the precursor ion *m/z* value and the collision energy for beam-type instruments (Figure 1).

As can be seen from Figure 1 and Figure 2, there is a great deal more information provided in a text format for the spectra in the nist\_msms database.

The MS/MS Database can be searched with any of the **Other Searches** or the **Incremental Name Search**. All that is necessary is to include the nist\_msms database in the **Include Libs** column of the **Search** dialog box for the **Other Searches** or as the single database to be searched in the **Incremental Name Search**. The **Unknown Spectrum Search** and **Structure Similarity Search** are also applicable. However, the user must understand that MS/MS spectra in general are different from EI spectra as well as the fact that structures in the MS/MS Database refer to the neutral form of the precursor ions.

If the nist\_msms database is not present, the **MSMS** tab is not displayed.

## Version 2.0g of the MS Search Program

The basic software of the **NIST MS Search Program for Windows Version 2.0** is unchanged. A number of new features have been added to the version 2.0g distributed with NIST 11.

- A new **Exact Mass Search** allows searching for exact mass of a precursor using its mass or exact *m/z* of a product ion. It takes into account 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 databases **mainlib**, **replib**, and **nist\_msms** using the **Any Peaks Search**.
  - Display of the exact mass with any spectrum that has an associated chemical formula.
  - Manage and search as many as 127 separate MS databases 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.
  - Maximum number of spectra in a database has been increased from 786,420 to 1,048,560.
  - Full compatibility with most recent NIST Peptide MS/MS Databases.
  - 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 11 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 to improve the reliability of the score when searching noisy MS/MS spectra.
  - Several convenience features.<sup>4</sup>

## Previous 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 databases, including the NIST MS/MS databases.
- Spectra of isomers and derivatives are displayed as replicate spectra.
- Tags (which can act as Field Names) can be entered into the "Comments" field of a user database 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 Database.
- 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.

---

<sup>4</sup> New convenience features added to version 2.0g include:

- 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.

---

## Tab Displays

**Version 2.0** of the **Program** uses tab displays. The tabs, **Lib. Search**, **Other Search**, **Names**, **Compare**, **Librarian**, and **MSMS**, are located at the bottom of the **Main Window**. Selecting a tab results in the display of a different Window. Each Window serves a different function. All six of these major tab displays have moveable divider bars (splitters), which can be re-orientated as well as tabs associated with various panes (windows) within a view. There are six such panes in the **Lib. Search** tab (Figure 3) and three in the **Other Search**, **Names (Incremental Name Search)**, **Compare**, **Librarian**, and **MSMS** tabs. The three left-side panes of **Lib. Search** tab and the single left side pane of the **Other Search**, **Names (Incremental Name Search)**, **Librarian**, and **MSMS** tabs are separated from the right-side panes by a vertical splitter bar (orientation of which is fixed; however, the position is adjustable). The **Compare** tab has three panes separated by two horizontal splitter bars. The panes on the left side of each Window are reserved for a list of spectra. Those on the right are for graphical presentations of an individual spectrum and for providing text information pertaining to that spectrum. The plot of the spectrum and the text information for that spectrum are separated by a horizontal moveable divider bar. The orientation for this divider bar can be changed to vertical and back by selecting the **Change Splitter Orientation** option when the Mouse pointer is in either of these two panes and the Right Mouse button is used to display the Right Mouse button (RMB) menu. 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 button) 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 whatever type search is desired except for the **User spectrum Search** and **Structure Similarity Search**. When a search type is selected from the **Search** menu in any tab display, the display changes to the appropriate tab (**Other Search** or **Names**) and displays the associated dialog box if one is required.

The **View** option on the **Main Menu** (top of Figure 3) allows the display of the **Tool Bar** and/or the **Status Bar** (bottom portion of Figure 3 containing **For Help press F1** and **Ident**). There are also options to turn off the display of the **Compare pane** 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**.

The pane on the top left of the **Lib. Search** tab is used for the **Spec List**, and the pane on the lower left of this Window is used for the **Hit List**. The pane on the left of the **Other Search** and **Names** tabs is used for the **Hit List**, and the pane 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 row or as graphic structures by selecting the **Names** or **Structures** tab at the bottom of the pane, except for the **MSMS** tab which does not have this alternate display option.

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

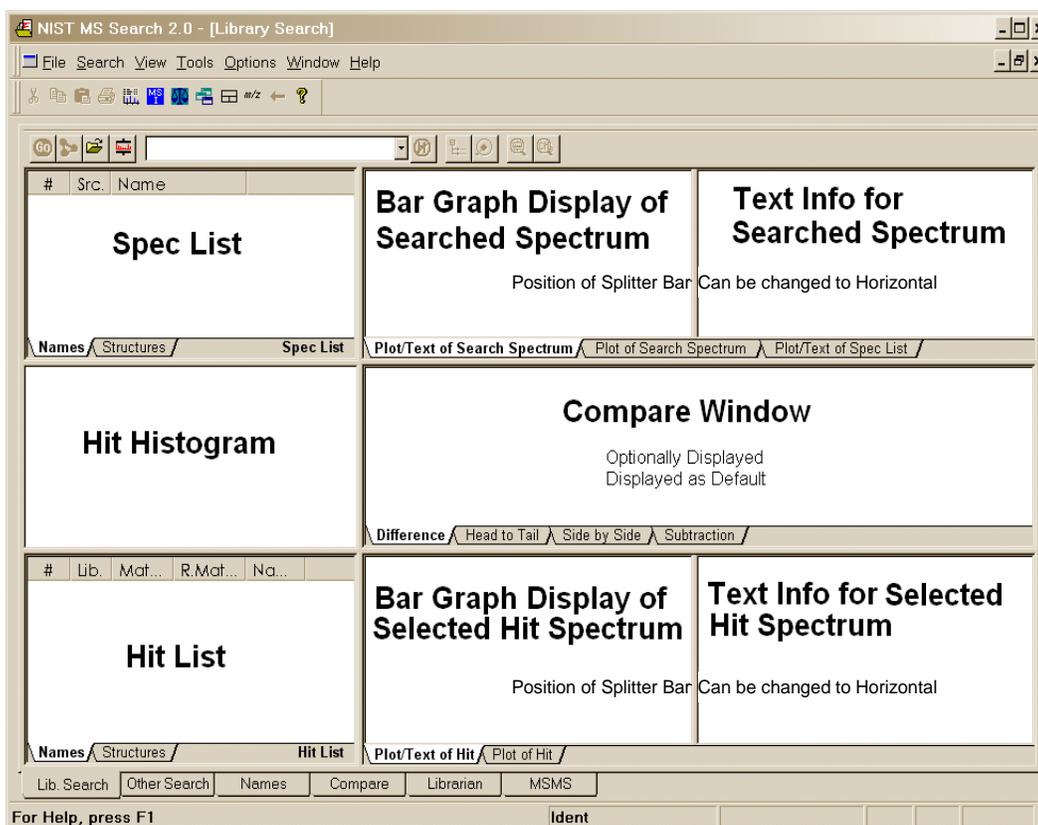


Figure 3. Lib. Search tab.

In some cases, these two **Hit Lists** will also have a *plus box* ☞ in front of the Hit number (only if the **Replicates On/Off** is set to Off). 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**, and some by **m** or **r**. If an **R** precedes the name, the spectrum is of the same compound as the Hit and is in the **NIST replib Database**. 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 4). Details are found by clicking on the **Help** button.

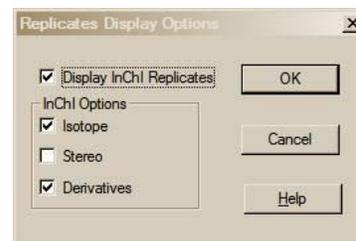
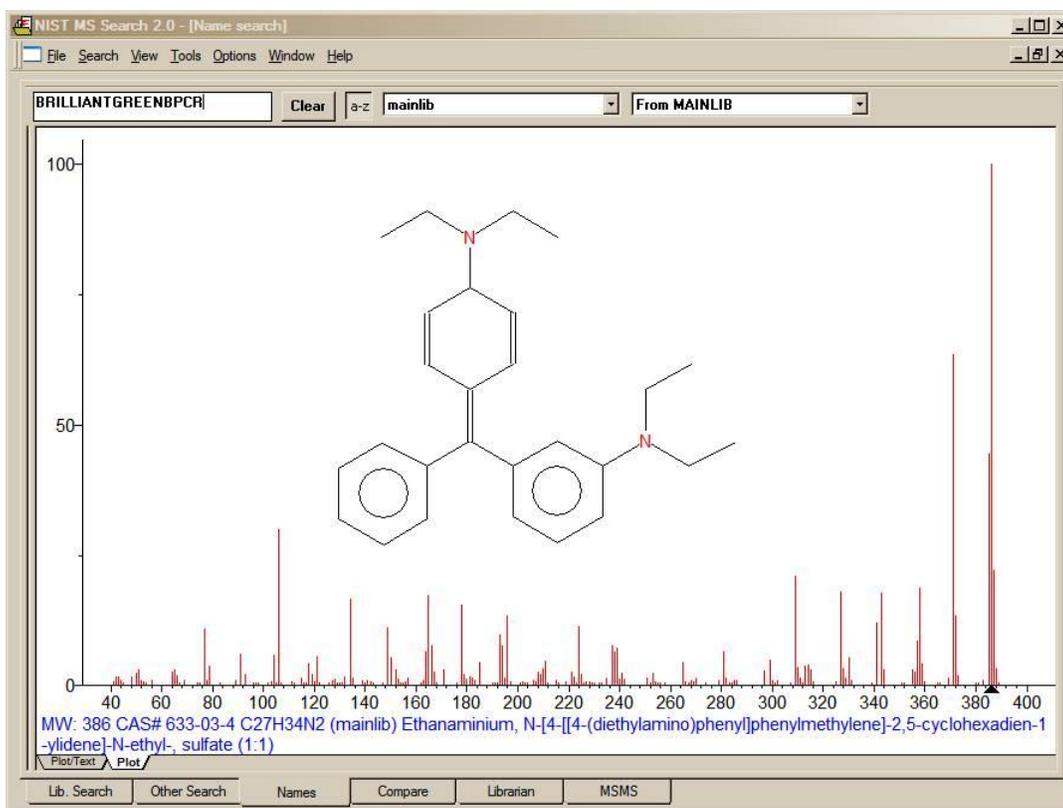


Figure 4

The **Spectral Plot** presentation pane (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** pane has the primary name associated with a spectrum (the **mainlib** name in the case of spectra in the NIST Database spectra that have a number of synonyms) preceded by the name of the database in parentheses [(mainlib) for the **NIST Main Database** or (mylib) for a user database], displayed under the bar-graph presentation. Optionally, the compound's Molecular Weight (nominal mass), CAS registry number, and Formula (elemental composition) can be displayed in this text string. This string can be forced to wrap so that all characters in the string will be visible (Figure 5). Also optional, a structure (if associated with the spectrum) can be displayed in the spectrum. If a structure has been requested and one is not associated with the spectrum, a colored ? will appear. 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 **B/Y 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 5.** The Spectral Plot display of a spectrum in the Name tab with the nominal mass, CAS registry number, elemental composition name of the database containing the spectrum, and the name associated with the spectrum. The display is wrapped, the structure is shown, and the peaks are not labeled.

The **Text Info** pane, 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** pane 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 panes are oriented by default with the **Plot** pane above the **Text Info** pane. Again, this orientation can be changed using the appropriate selection on the RMB menu when the Mouse pointer is on either of the two panes. 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 database; 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 DB archive); ID number (position in the database containing the spectrum); name of the database 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 Database)/Comments (user databases); 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. If a method is being sought for the analysis of the displayed compound in a specific matrix, or using a certain column type, the **Find** selection (new to v.2.0g) on the RMB menu when displayed in a **Text Info** pane can be very useful, especially because each record not only contains the specifics about the GC column but also the title of the journal article from which it was taken (Figure 6).

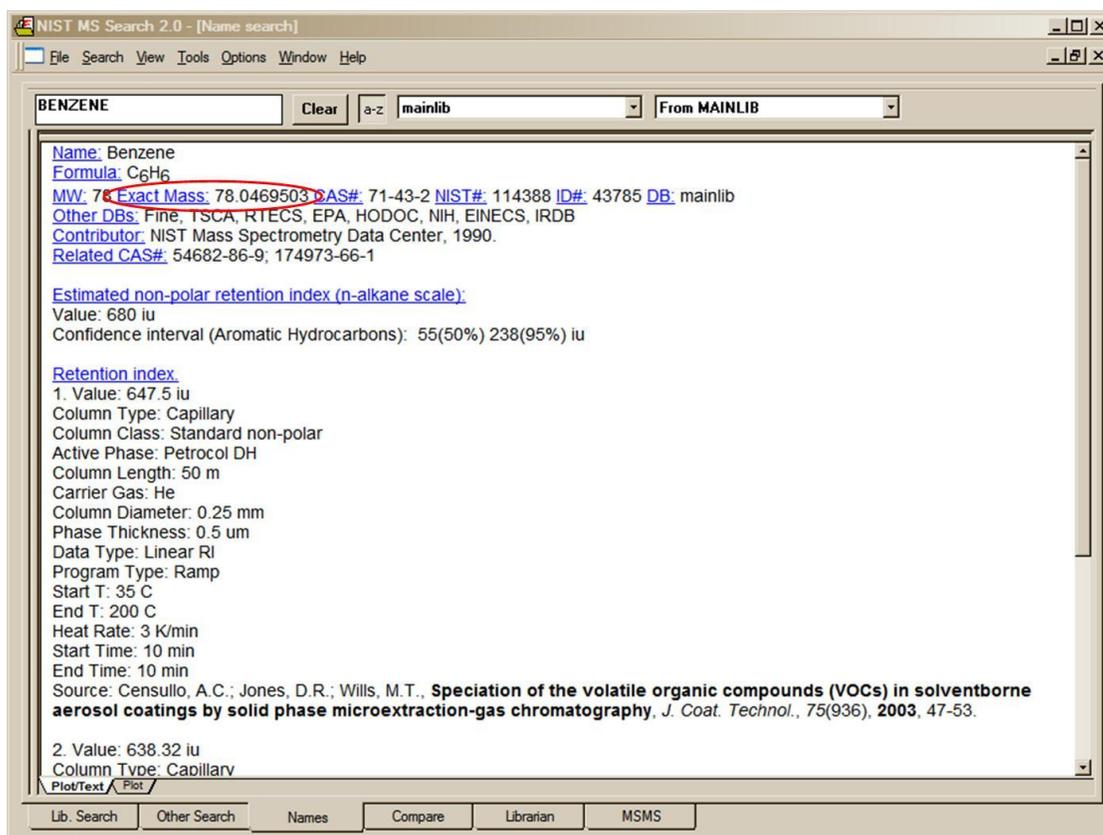


Figure 6. The Text Info display with Compound Information and Retention Indices as the only selected display options. Note the detail of the GC Method and the title of the journal article. Note the display of the Exact Mass value in the Header. This is new in v.2.0g and appears for any spectrum with an associated elemental composition.

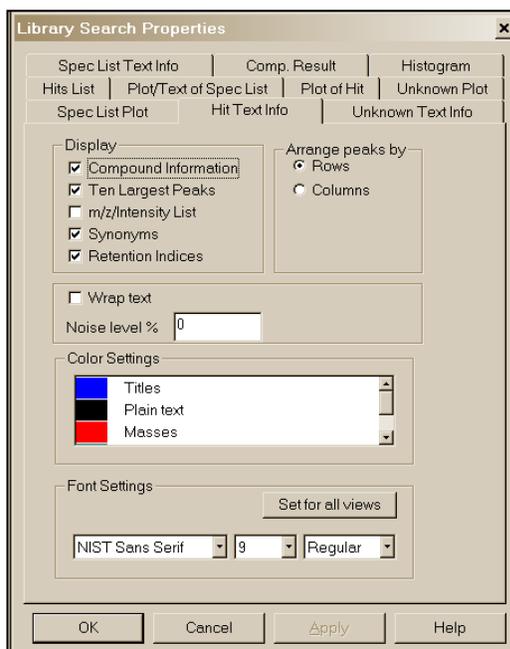


Figure 7. Properties dialog box displayed by selecting Properties on the RMB menu displayed when the Mouse pointer is on one of the panes of the Lib. Search tab.

It should be noted that the GC Methods/RI Database is also available separate from NIST 11 with its own Search program, which allows for more extensive use of this Database.

The RMB menu selection **Properties** will result in the display of a multiple-tab dialog box (Figure 7). 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 pane 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 databases to be searched.

Just below the **Tool Bar**, if displayed (below the Main Menu if the **Tool Bar** is not displayed), in the **Lib. Search** tab view (Figure 3) is the **Button Bar**. An explanation of this **Button Bar** is provided in Figure 8.

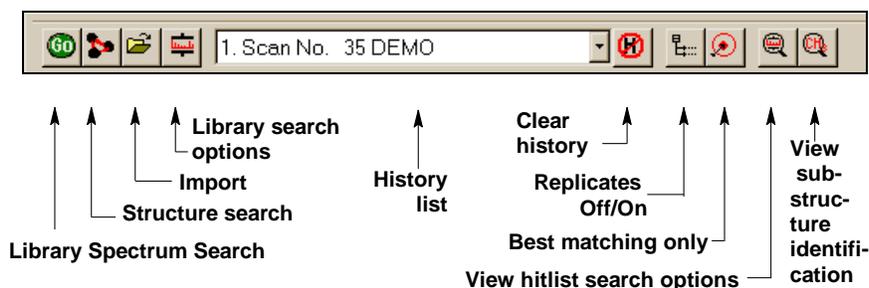


Figure 8. Lib. Search's Button Bar.



Figure 9. List of databases used in a User spectrum Search along with the Search Options.

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** *checkbox* 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. **NOTE:** MS Search v.2.0d used different index files for the **Structure Search** in user databases than previous versions; therefore, files created before v.2.0d should be rebuilt using **Rebuild Structure Search Database** from the **Tools** menu. This should also be done for each new user database connected to the **MS Search Program** and every time spectra with structure are added, removed, or edited in a user database because the structure search index files are not updated automatically.

The **Go** button will launch a search of the highlighted spectrum in the **Spec List**. The button to the right of the **Go** button will launch a **Structure Search** of the highlighted structure in the **Spec List**. The next button to the right (with the file folder icon) is used to import spectra from a text file. The last button, just before the drop-down list box, is used to display the **Library Search Options** dialog box. This dialog box is used in selecting databases 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

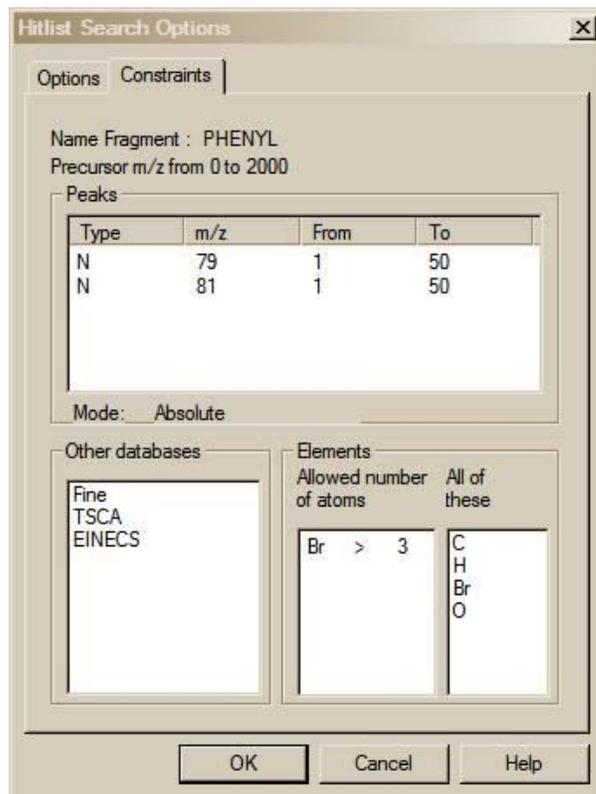


Figure 10. List of Constraints used for a specific search along with their values.

---

The first button to the right of the **Search History** list box is used to clear the search history (**Clear History** button). The next button (the **Replicates On/Off** button) to the right is used to toggle the listing of replicate spectra on and off when the **mainlib Database** of the **NIST/EPA/NIH Database** is searched.

The button to the right of the **Replicates** button is used to display only Hits with the highest Match Factor when more than one Hit has the same CASrn. This is the **Best matching only** button. This feature is especially useful when more than one database is being searched, and some of the spectra in these databases 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.

The button to the right of the **Best matching only** button is the **View hit list search options** button. Selecting **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 databases searched (and type of Search and Search Options for **User spectrum Searches** only). The **Constraints** tab (if present) lists all different constraints along with values used in the search (Figure 9 and Figure 10).

It should be noted that even though a **User spectrum Search** is carried out against a list of databases that does not include **mainlib** and **replib** (see Figure 27), these two databases will be listed as having been included as long as the **Display InChI Replicates check box** is selected in the **Replicates Display Options** dialog box (see Figure 4). It does not matter that none of **InChI Options** have been selected. This will result in the display of the symbol indicating replicates were found when the user-only database was searched. This means that **InChI** search options are available for those spectra in databases like the Wiley Registry of Mass Spectra and various other databases provided by Wiley and Others in the NIST format, which have CASrn present in **mainlib**.

#### **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; therefore, a slider bar on the right side of the **Constraints** Window in the **Search** dialog box is needed. 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 cleared. 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. If a Constraint is selected and is not visible in the Window, a bad result may occur. A **Clear All** button has been added to the **Constraints** tab of the **Search** dialog box. 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. The **Clear All** button removes the check from the *check boxes* of all the constraints and the check in the **Use Constraints check box**. Any previously entered values will be retained and displayed when that Constraint is selected so that the previous values can be used or replaced. 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 **Neutral Loss, Hybrid Neutral Loss 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 EI Database 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 **User 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 **User 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 database(s) being searched.

There are five fields on the right side of the **Status Bar**. The left-most of these five fields will have displayed the selected search type in the "Spectrum Search Type" area of the **Search** tab of the **Library Search Options** dialog box (**Neutral Loss**, **Hybrid**, or **Simple** if the **Similarity Search** is selected; and **Quick** or **Ident** depending on whether a **Quick** or **Normal Identity Search** has been selected).

The next four fields to the right of the box indicating the selected search type pertain to the currently displayed search results. The first box to the right contains the type of search that was used in the displayed search results. The next box to the right will have contained "Rev" if the **Hit List** is sorted according to the Reverse Match value (the Match Factor obtained by ignoring all peaks that are in the sample spectrum but not the database spectrum). To set this mode, select **Reverse Search** *check box* in the "Spectrum Search Options" section of the **Search** tab of the **Library Search Options** dialog box. The third box will contain the number used as a precursor MW value in a **Neutral Loss** or **Hybrid Similarity Search**. If an **Identity Search** has been carried out, the fourth box will contain "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 11) 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; the **Isotope Calculator**, **MS Interpreter**, and **AMDIS** also may be launched from the **Tools** menu.

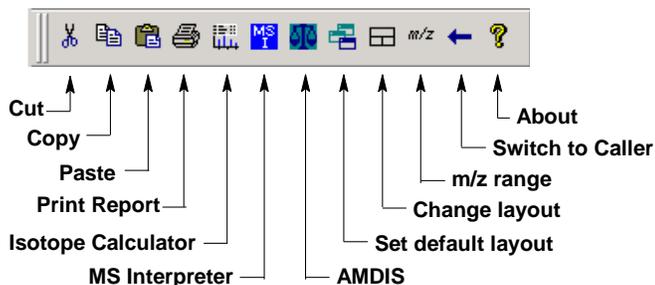


Figure 11. NIST MS Search Tool Bar.

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 12) 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.

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 this dialog box have to do with labeling spectra normally or with neutral loss values. Normal display is the most common.

## Neutral Loss Display

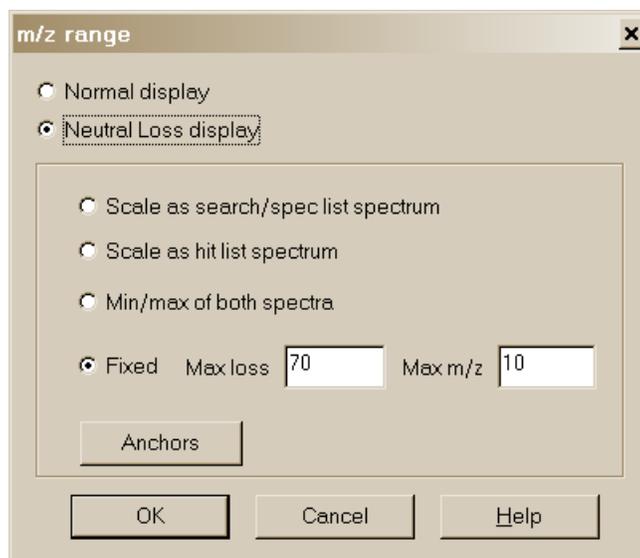


Figure 12. *m/z* range (for Spectra Display) dialog box.

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 13.

A feature of the MS Search v.2.0g is the ability to display a spectrum with neutral loss labeling. Even after a search has been performed, the sample spectrum and the database spectra may be labeled as neutral losses with the first label on the right-most peak being 0 (Figure 14). 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).

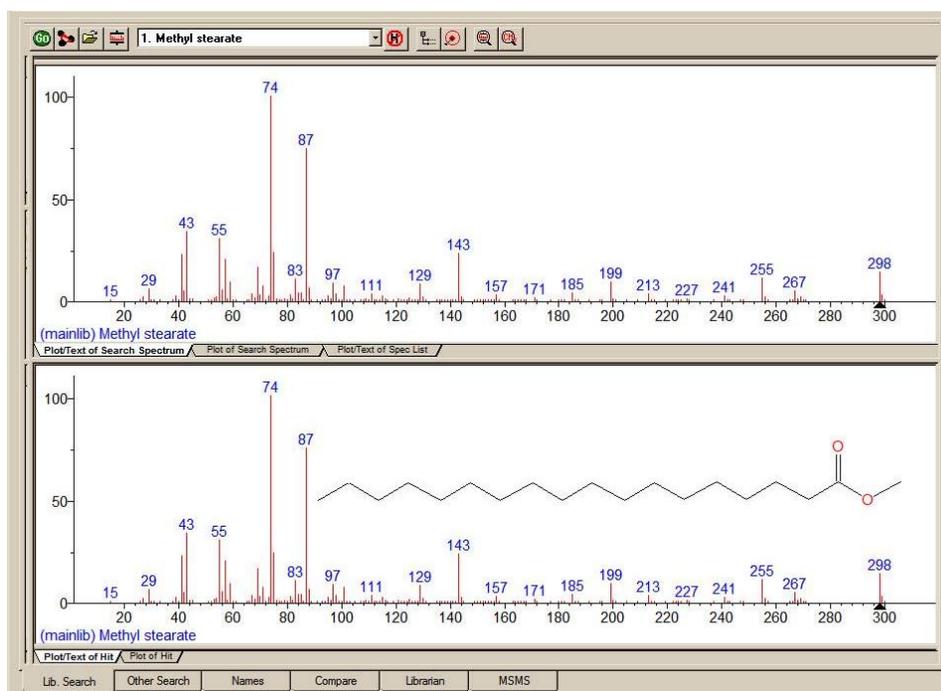


Figure 13. Nominal spectrum labeling where the top spectrum is of an unknown compound and bottom spectrum is of the first Hit from a search of the NIST/EPA/NIH Mass Spectral Database (NIST 11).

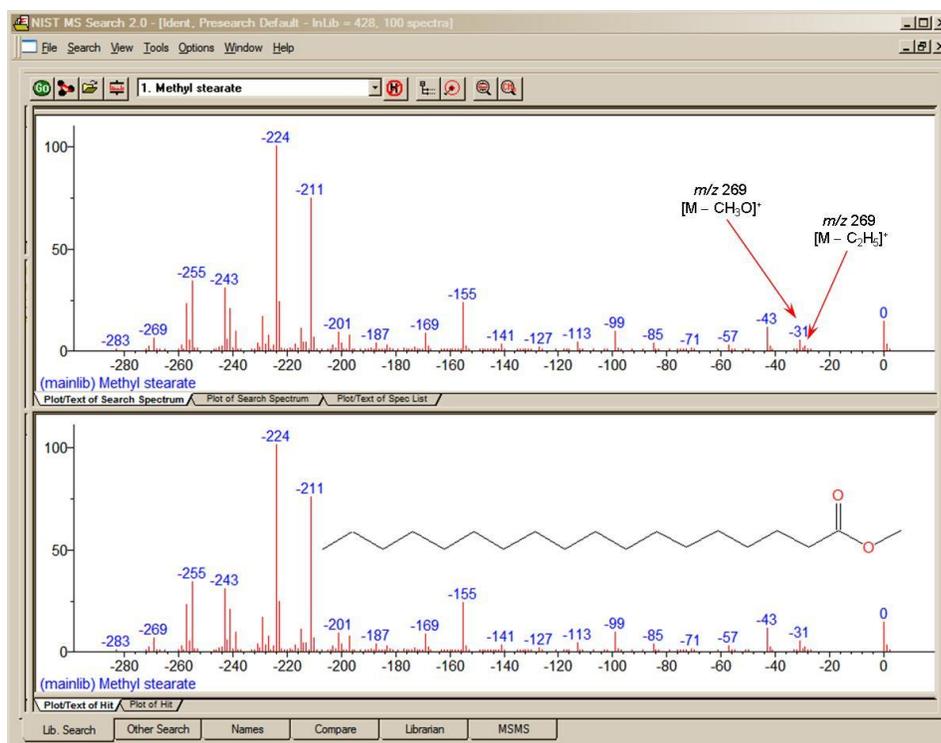


Figure 14. Neutral Loss spectrum labeling where the top spectrum is of an unknown compound and bottom spectrum is of the first Hit from a search of the NIST/EPA/NIH Mass Spectral Database (NIST 11).

## 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 databases: a user database and/or the **NIST 11 Databases** (Figure 3). The spectrum and structure searches can be done against multiple databases (maximum of 127), and the selected databases 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 pane 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 pane 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 pane contains the **Hit List** at the bottom. By default, the columns for the number of synonyms and presence in other databases are not displayed.

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 panes. By selecting **View** from the **Main Menu Bar**, the display of the **Compare** pane can be deselected. If the Mouse pointer is in the **Search Spectrum/Spec List Spectrum** Window or the **Hit List Spectrum** pane, the RMB 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.

## The Other Search Tab

The second tab display, **Other Search** (Figure 15), is used to perform a **MW (Nominal Mass)**, **Exact Mass**, **Formula** (elemental composition), **Any Peaks**, **Sequential Method**, **CAS registry number**, **NIST Number**, or **ID Number Search**. The **NIST Number Search** can only be carried out on the **NIST Main Database (mainlib)** and the **Replicates Database (replib)**. **NIST Numbers** establish correspondence between spectra in different releases of the **NIST/EPA/NIH MS Database**. The **ID Number Search** can only be carried out on one database at a time. The other options in this group can search multiple databases (maximum of 127), and different sets of databases can be specified for each different type of search.

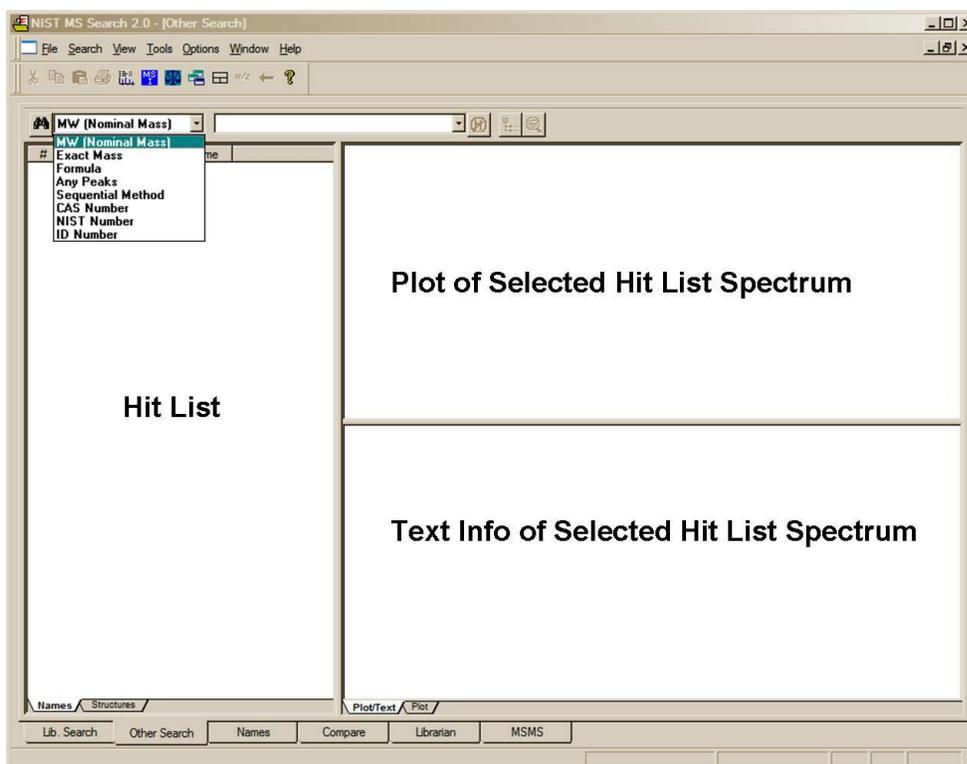


Figure 15. Other Search tab.

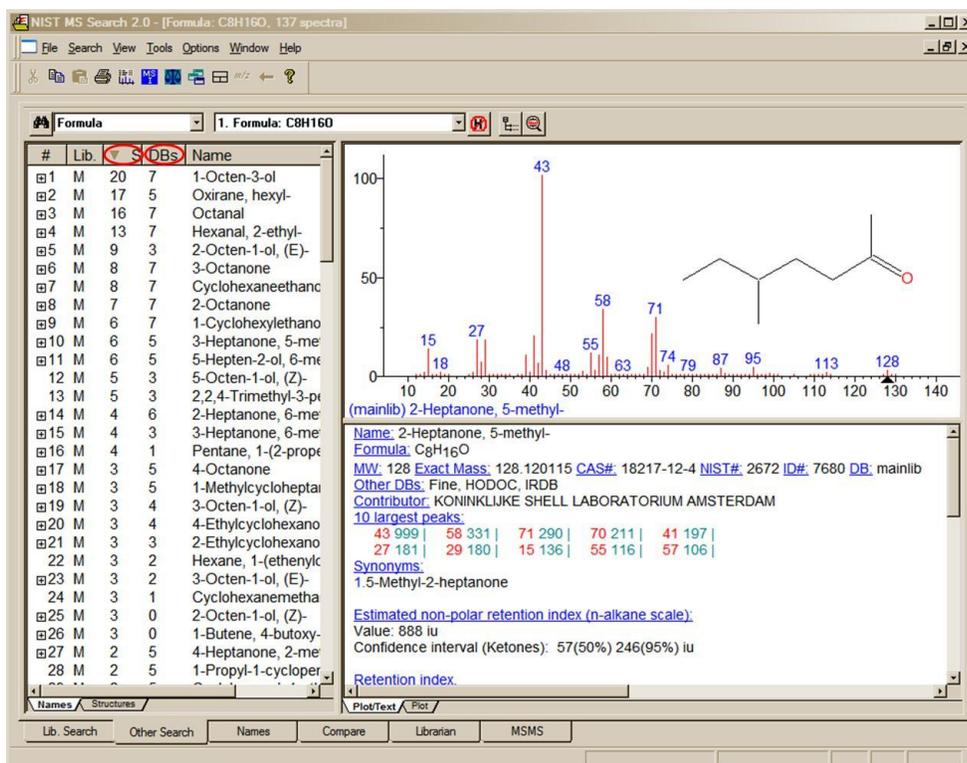


Figure 16. Results of a Formula Search showing the number of “synonyms” for the Hits and the number of other “databases” in which the Hits can be found columns.

As shown in Figure 16, another new feature in v.2.0g of the **Program** is the display of columns showing the number of synonyms and the number of other databases in which the Hit can be found. Not only are these two 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** pane.

## The Name Tab

The third tab, **Names**, is used for the NIST signature **Incremental Name Search (INS)** (Figure 17). The **INS** can only be performed against one mass spectral database at a time. The database to be searched is selected by use of the **database select** 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.

The drop-down list box just to the right of the **a-z** button is used to select the database to be searched. The second list box is used to display replicate spectra for the **NIST 11 mainlib Database**. 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 Database**. **\*Replicate** identifies a spectrum that is of a heavy isotope variant of the compound, a stereoisomer of the compound or a derivative of the compound. The sequence number has no meaning. All spectra that are from the **NIST Replicates Database** appear at the top of the list.

One factor that makes the **Name Search** used with the **NIST/EPA/NIH Mass Spectral Database** so powerful is the fact that many of the spectra in the Database have multiple synonyms. A spectrum can be retrieved by any one of these names. 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 17).

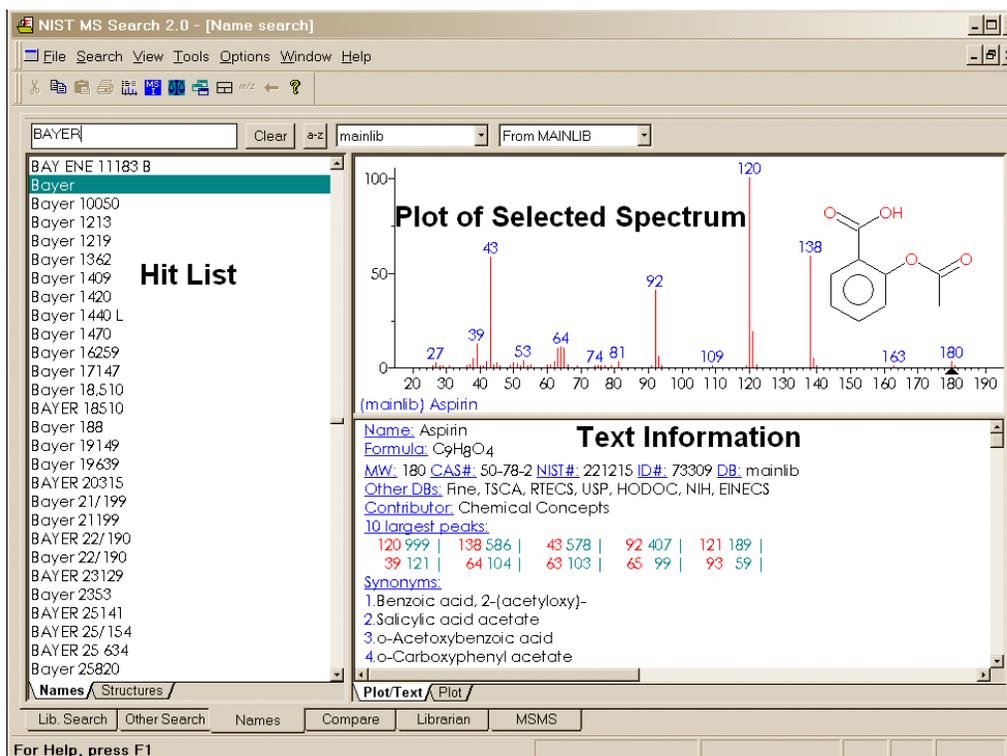


Figure 17. Incremental Name Search tab.

## The Compare Tab

The fourth tab, **Compare** (Figure 18), 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** pane (**Difference**, **Head-to-Tail**, **Side-by-Side**, **Subtraction**).

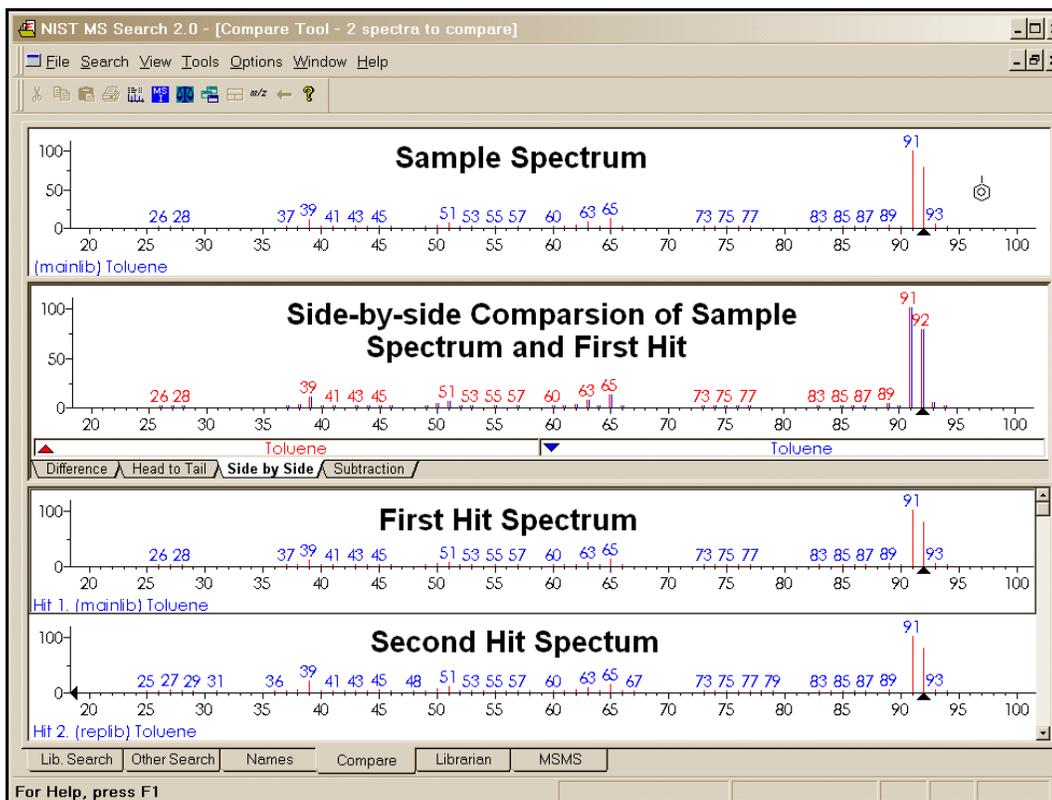


Figure 18. Compare tab.

## The Librarian Tab

The fifth tab, **Librarian** (not shown), is used to deal with spectra that are in user databases or that are to be added to user databases. The layout of this tab is the same as the tab used for the **Incremental Name Search** with the exception that the far left pane contains the **Spec List** rather than a **Hit List**. This is the same **Spec List** as displayed in the upper left pane 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.

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

**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 Databases** and **MS/MS Searching**.

---

# Spectral Analysis Utilities for NIST 11

## ***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 database 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. In order 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 databases (environmental, flavor and fragrance, and drugs and toxins) that were derived from the NIST 02 MS Database. **AMDIS** has a range of other features including the ability to search the entire **NIST 11 MS Database** 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 databases. 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 DATABASES** 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** pane 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 20). **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** pane and the **Spectrum** pane 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, two files are created: one containing the spectrum (\*.MSP) and the other a structure (\*.STB).

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 group. It also has a simple 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 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).

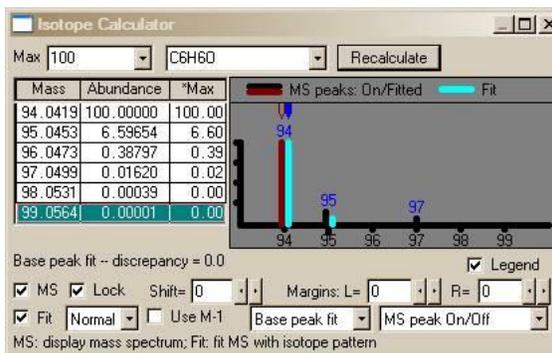


Figure 19. View of MS Interpreter's isotope calculator used to compare observed abundances to theoretical values.

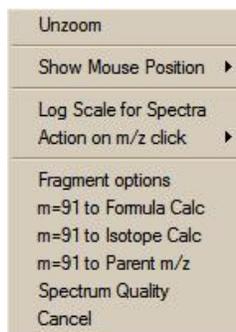


Figure 21. Right Mouse button menu displayed when the Right Mouse button is on the Plot portion of the MSI Program.

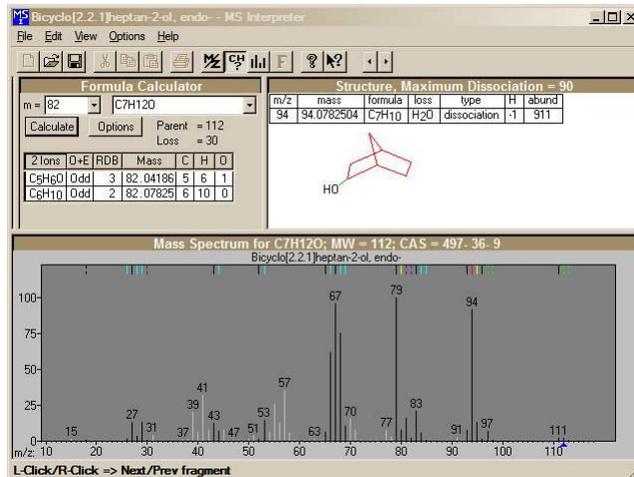


Figure 20. The main Window of MS Interpreter. The black lines indicate the peaks in the mass spectrum that the MS Interpreter can explain. The white lines indicate the ones it cannot. The color coding is settable by the user to allow specific types of ion formation mechanisms to be shown.

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 22).

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** panes.

NIST11 version of MS Interpreter has the following new features:

#### 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**.

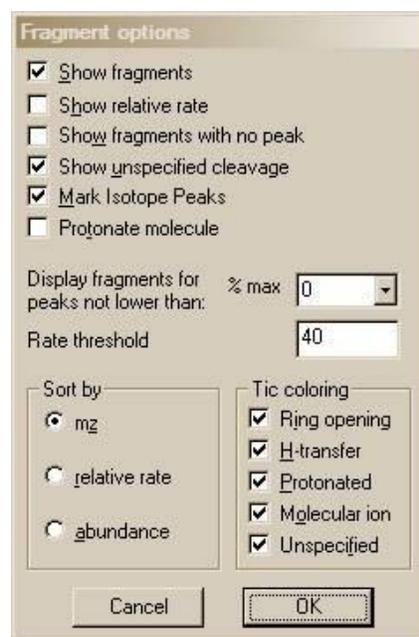


Figure 22. Fragment options dialog box.

#### 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;
- 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:\nist11\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.

## Example Searches

### User 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 or from a text file (\*.MSP). A spectrum can also 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 **User 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** pane or **Text Info** pane in any tab, or in a **Compare** pane; display the RMB menu and **click with the Right Mouse button any spectrum** in a List of spectra or Subtraction spectrum in the **Compare** pane; 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 23). 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 24) 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 25). This is where information about the *m/z* values of peaks is controlled.

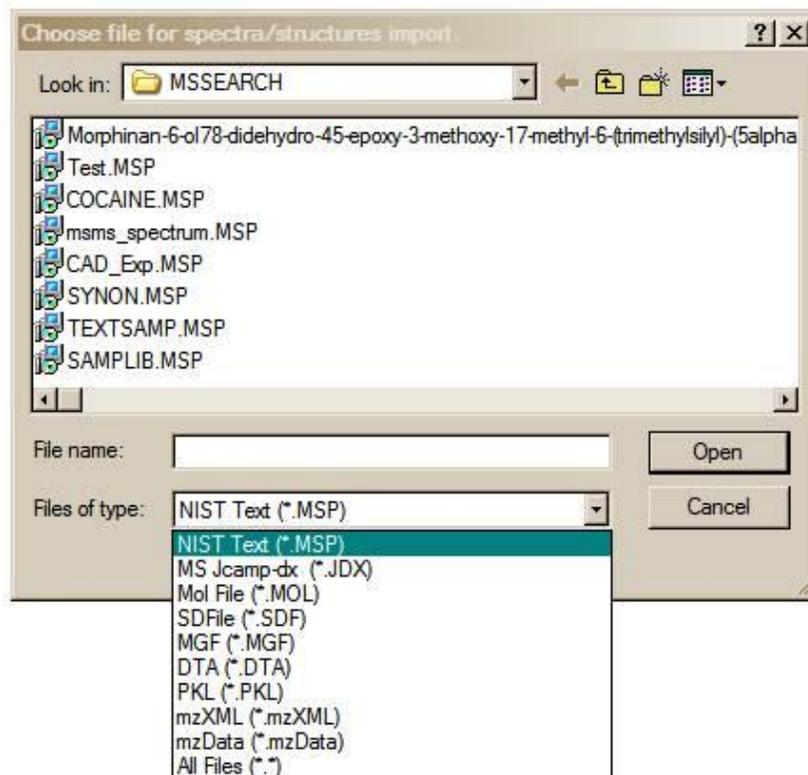
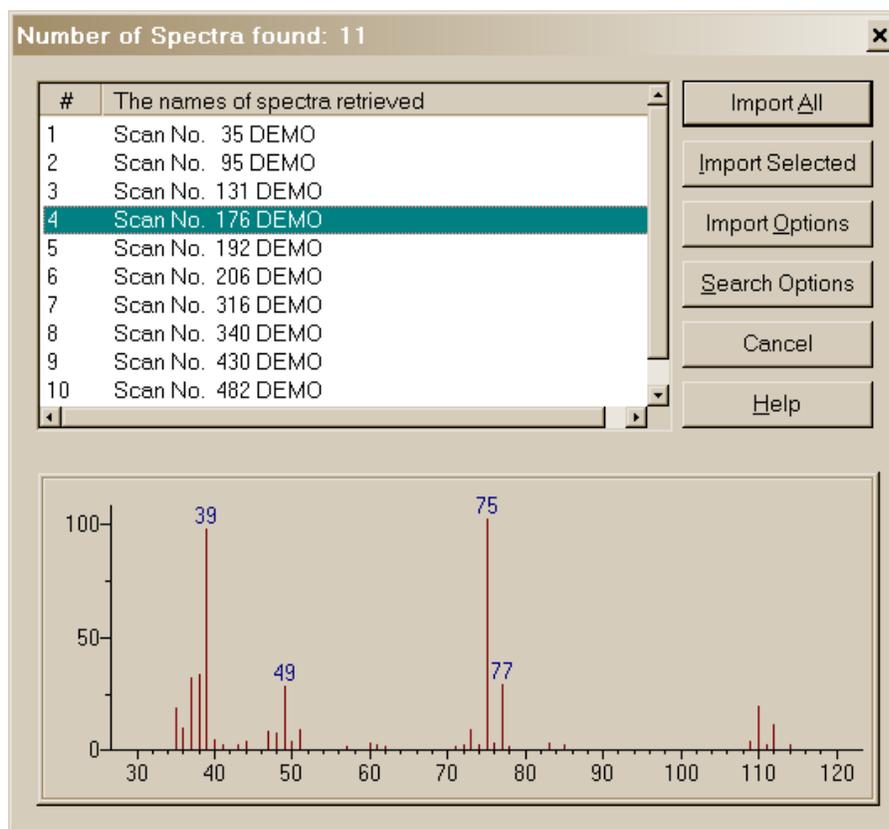
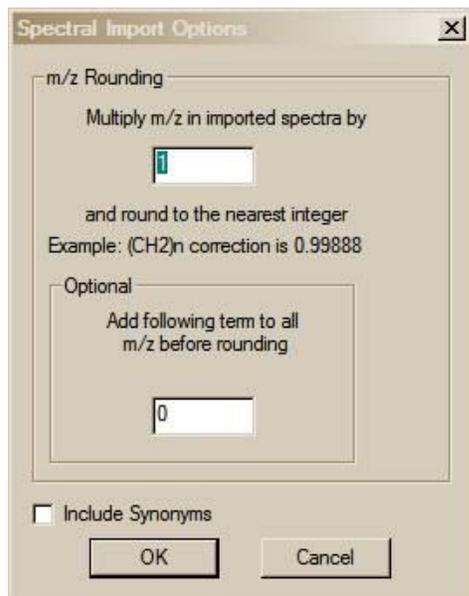


Figure 23. File Selection dialog box with some of the available file types. MSP and JDX files are spectra, and MOL files are structures.



**Figure 24. Spectrum Import dialog box with a single spectrum from a multiple spectrum file highlighted and displayed in a bar-graph format.**

The two fields in 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 11 MS Database**. 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 in order to be properly searched against the **NIST 11 MS Database** or another database that contains only integer  $m/z$  values.



**Figure 25. Spectral Import Options. Imported spectra integer  $m/z$  and no mass defect correction.**

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 will have to 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".

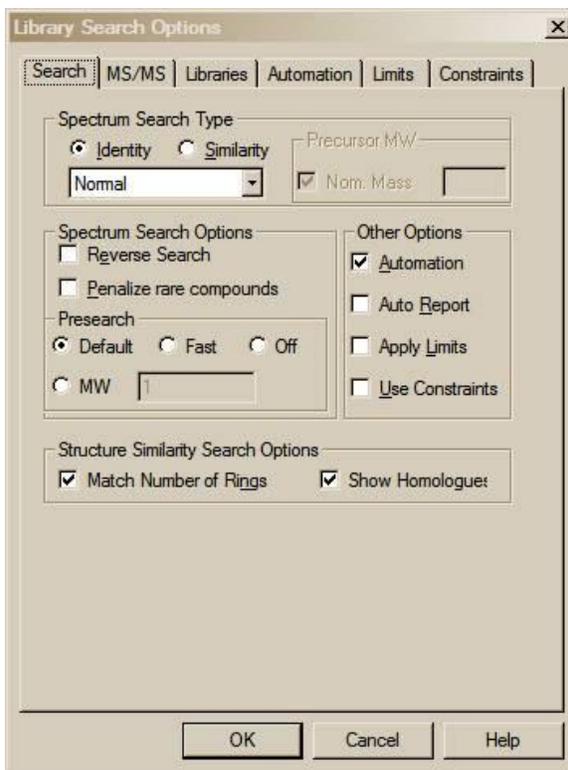


Figure 26. Library Search Options Search tab for an “Identity” Normal search.

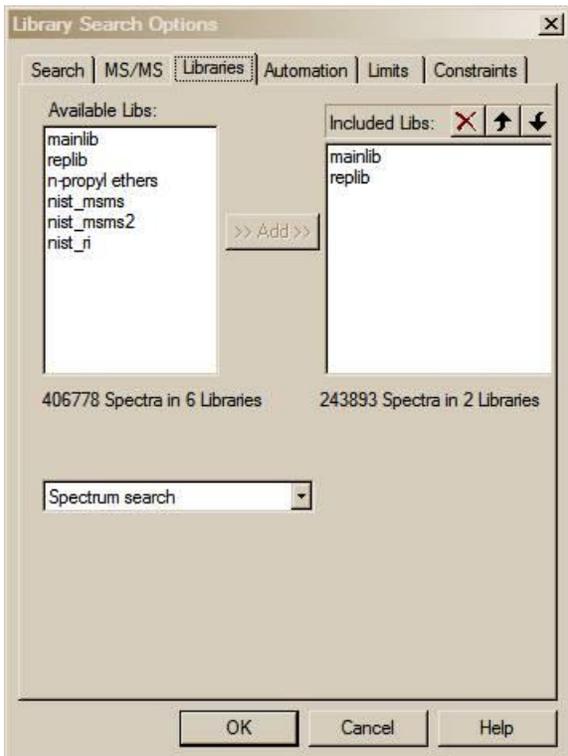


Figure 27. Library Search Options Libraries tab.

## Library Search Options

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 26).

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 three options) and **Similarity** (with four options). These different searches and their options are explained in the Help Screens (topic Spectrum Search Type).

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.

## MS/MS

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

## Libraries

The **Libraries** tab (Figure 27) of this dialog box is used to select the databases to be included in the **User spectrum Search**. As many 127 databases can be searched at the same time. This includes the **NIST 11 Main Database (mainlib)** and the **NIST 11 Replicates Database (replib)** and multiple user databases as well as other commercially available databases from sources other than NIST that are in the **NIST MS Search** database format. By default, three other NIST 11 Databases are also installed: the **nist\_ri**, **nist\_msms**, and **nist\_msms2**. This means that there is space to accommodate an additional 122 user databases. If more MS databases need to be accommodated, the less used ones can be moved to another folder. To add databases to be searched in the **Available Libs** pane of the dialog box, highlight them and click on the **Add** button. To exclude databases, highlight them in the **Included Libs** pane and then click on the **X** button above the upper right portion of this pane. The order in which databases are searched can be changed by highlighting the database to be moved in the **Included Libs** pane and clicking on the Up or Down arrows in the upper right portion of this pane. Each click will move the database one position. The order in which databases are searched is not important because the **Hit List** has the 100 Highest Match Factor Hits regardless of what database they came from.

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

## Automations

When a **User 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 26 and Figure 28).

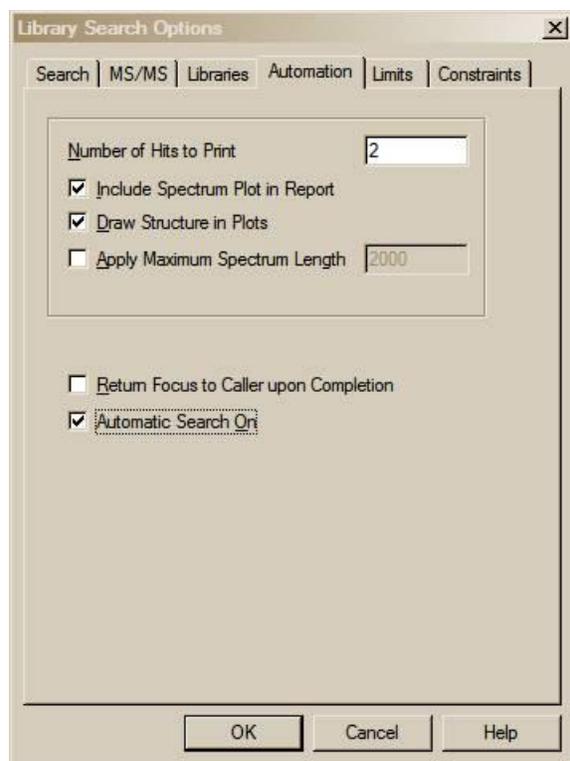


Figure 28. Library Search Options Automation tab.

*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 **User 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 **User 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 databases, 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 **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 26: it turns on automatic searching. Selecting or deselecting this *check box* on either tabs toggles the selection/deselection of the *check box* on the other tab.

Search results can automatically be printed whether or not the search is automatically launched. *Check box Auto Report* (Figure 26) turns on this feature. "Number of Hits to Print" field is located 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

## 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 databases of compounds having specified (nominal mass) molecular weight (excludes replib).	<b>Other Search</b>
<b>Exact mass</b>	Finds spectra in selected databases of compounds having an exact mass corresponding to specified accurate mass with a specified variance (excludes replib).	<b>Other Search</b>
<b>Formula (Elemental Composition)</b>	Finds spectra in selected databases of compounds with specified formula (excludes replib).	<b>Other Search</b>
<b>Any Peaks</b>	Finds spectra in selected databases based on user-specified peaks.	<b>Other Search</b>
<b>Sequential Method</b>	Searches selected databases 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 database 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 <b>mainlib</b> and <b>replib</b> 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 (with the exception of 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 30 and 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 31), which allows for the selection of the databases 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

The **Any Peaks Search** became very popular when it was first developed over 35 years ago for use with the on-line Mass Spectral Search System from, first, NIH and then the combined EPA/NIH system. This has evolved into the powerful tool that exists in the **NIST Mass Spectral Search Program**. This has been made even more powerful by a *New* feature in v.2.0g, which is the addition of the ability to use accurate mass values with variance limits to search the **NIST/EPA/NIH Mass Spectral Databases**.

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 82 and press <Enter>*. This moves the cursor to the "From" field. *Type 70 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.

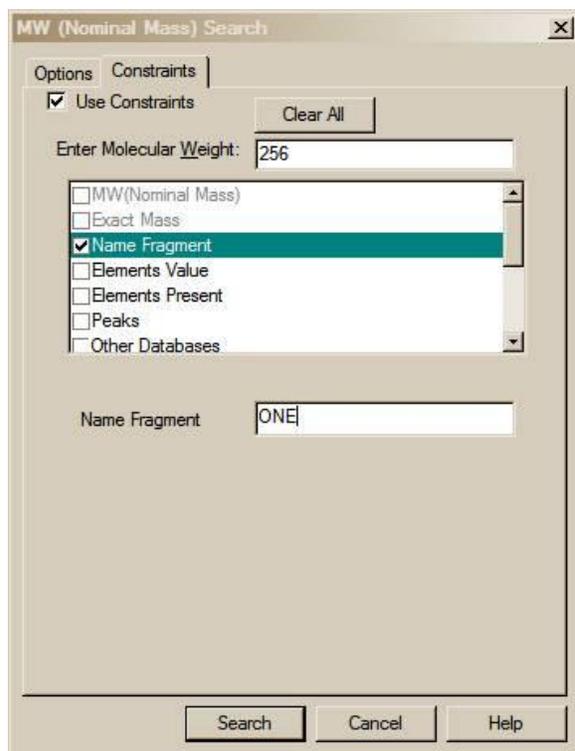


Figure 29. MW (Nominal Mass) Search dialog box with Use Constraints selected and a Name Fragment entered.

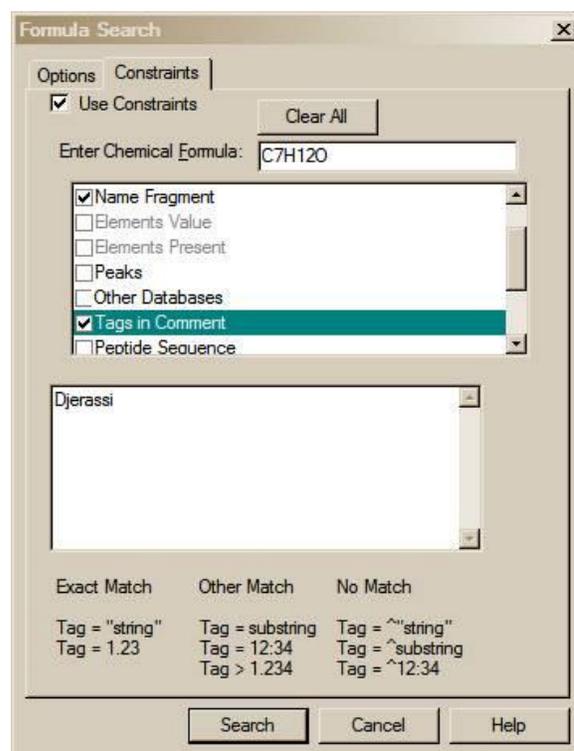


Figure 30. Formula Search dialog box with Use Constraints selected and a Tag in Constraints entered to yield only spectra contributed by Carl Djerassi.

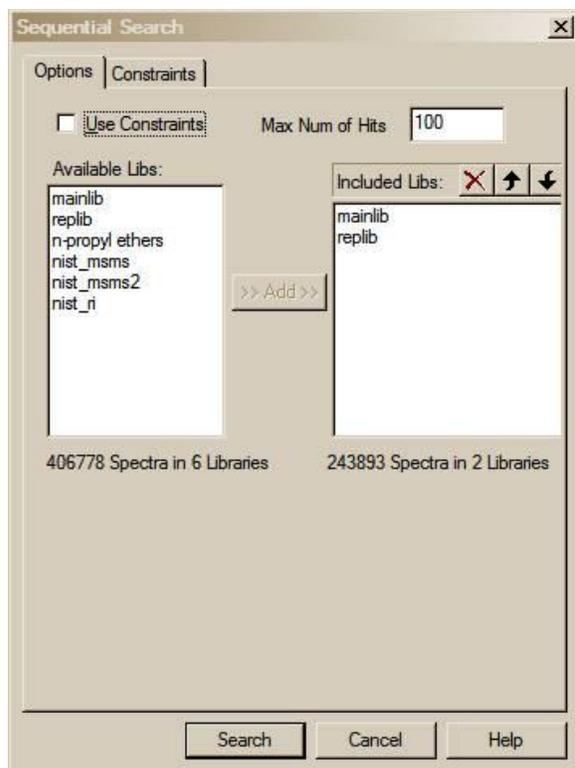


Figure 31. Options tab of the Sequential Method Search dialog box showing that a maximum of 100 Hits will be displayed.

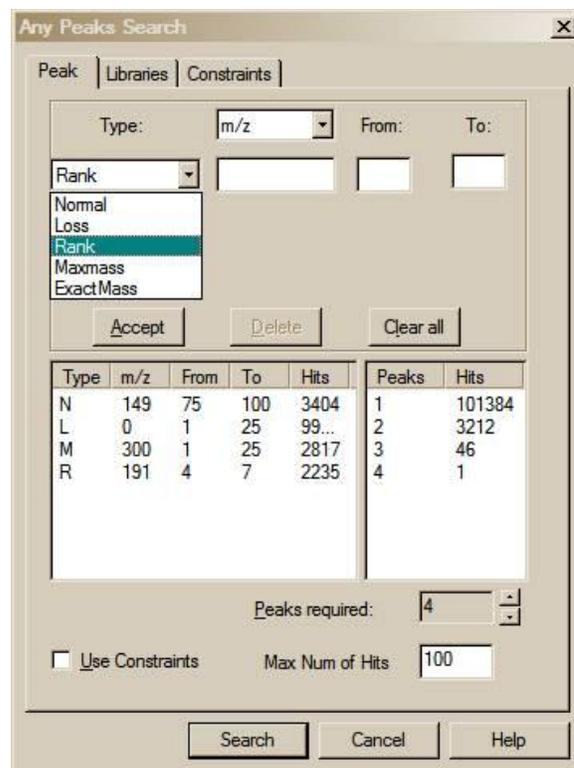
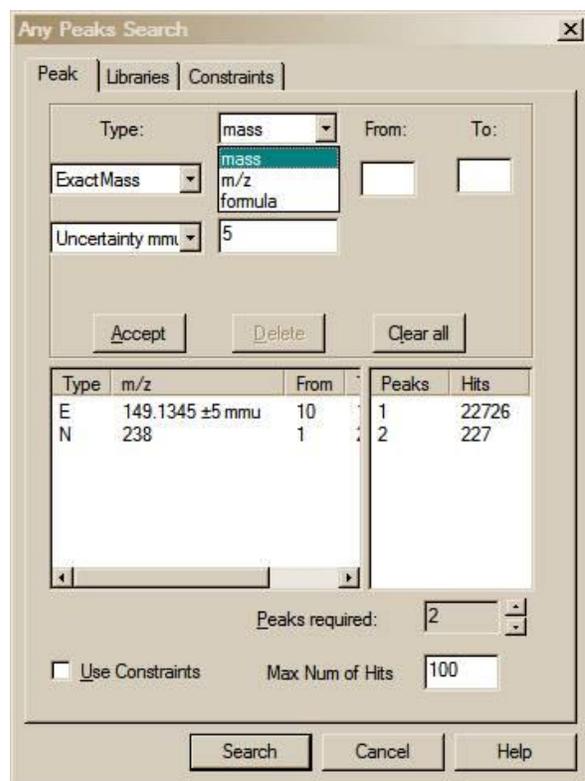


Figure 32. Peaks tab of the Any Peaks Search dialog box showing the selection of the four different types of peaks and intensity ranges.



**Figure 33. Any Peaks Search dialog box, showing an entry of an accurate mass value with variances. Note the position of the slider bar.**

(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 database. 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 "Formula" 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. In an **Any Peaks Search** of the **NIST 11 mainlib**, using the "Exact Mass" Peak type and a formula entry, with the intensity range of 10–50%, 1 peak is found in 5,762 spectra and 2 peaks are found in 1,323.

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-positives. 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!).

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 31 and press <Enter>. Type 5 and press <Enter>. Type 25 and press <Enter>.* The retrieved spectra must have a peak at 31  $m/z$  units lower than the molecular peak that exhibits intensity between 5% and 25% of full scale. This reduces the number of spectra in the **NIST/EPA/NIH Mass Spectral Main and Replicates Databases** that match the entered criteria from 1796 for the first peak to 122 that have both peaks.

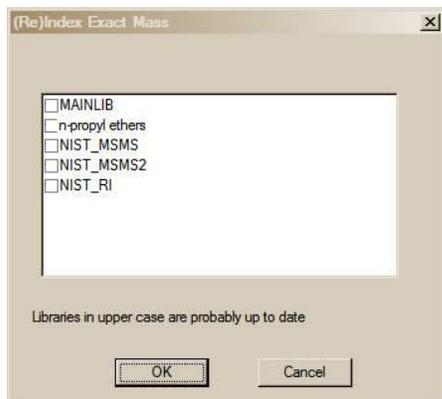
The next peak will be a "Maxmass" peak. *Click on the ↓ on the right side of the "Type" field. Click on "Maxmass". Type 303 and press <Enter>. Type 1 and press <Enter>. Type 50 and press <Enter>.* There will only be eight spectra that contain all three peaks of the designated type and over the designated intensity range.

Select the **Search** button in the **Any Peaks Search** dialog box to copy the spectra found into the **Hit List**.

Figure 33 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

## Exact Mass Search

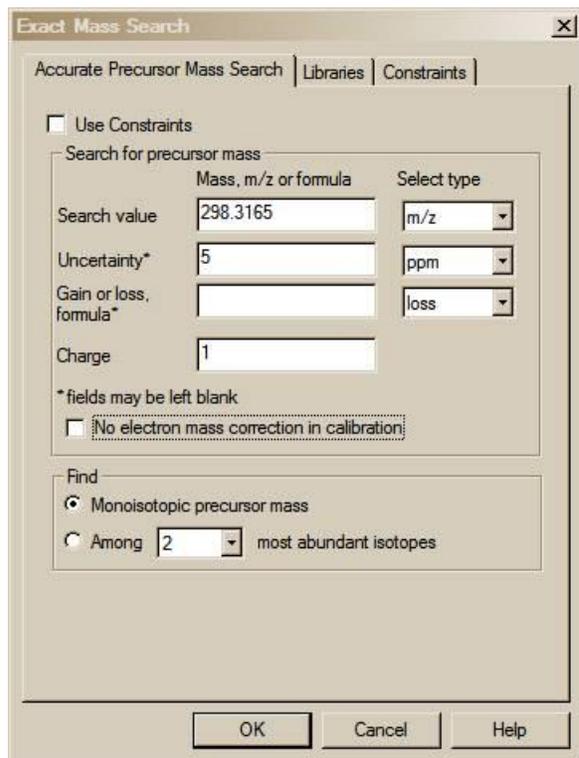
New in v.2.0g of the **NIST MS Search Program** is the ability to perform an **Exact Mass Search** and use of **Exact Mass** (alternately with nominal mass) as a Constraint in **User Spectra Searches (Lib. Search tab)**, **Any Peaks Search**, and **Sequential Search (Other Search tab)**. Such a search is not restricted to the **NIST/EPA/NIH**



**Figure 34.** Dialog box to allow for indexing databases according to exact mass.

**Databases**; however, the non-NIST database must first be indexed according to exact mass. Select **Tools** from the **Main Menu** and then select **(Re)Index Exact Mass**. This will result in a dialog box being displayed (Figure 34). The time to index a database will depend on the size of the database. All **NIST 11 Databases** are already indexed according to exact mass.

After the desired databases 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. 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 35 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 of 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 databases to be searched. Like the **MW (Nominal Mass) Search** and the **Formula Search**, the **NIST 11 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. The **Exact Mass Search** is very useful for accurate mass data that is generated by TOF, Orbitrap, and other instruments capable of proving  $m/z$  value in the millimass unit range. When GC/MS data are acquired with such an instrument, the accurate mass constraint can be the concluding information required for an unambiguous identification.



**Figure 35.** Exact Mass Search's dialog box.

## Search by Name

A useful feature of the **Program** is its ability to search through hundreds of thousands of names to find mass spectra. To see if a mass spectrum of a compound whose common name is provided is in the **NIST/EPA/NIH Mass Spectral Database (NIST 11)** of approximately 213,000 compounds, begin typing the name until it appears in the **Hit List** pane of the **Name Search** tab. Highlight the desired name, and the display in the **Plot/Text** and **Plot** Windows will be refreshed with the appropriate displays 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** pane 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 **Past** selection.

## Sequential Method

This method allows every spectrum in the selected databases to be subjected to a created search criteria. 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.

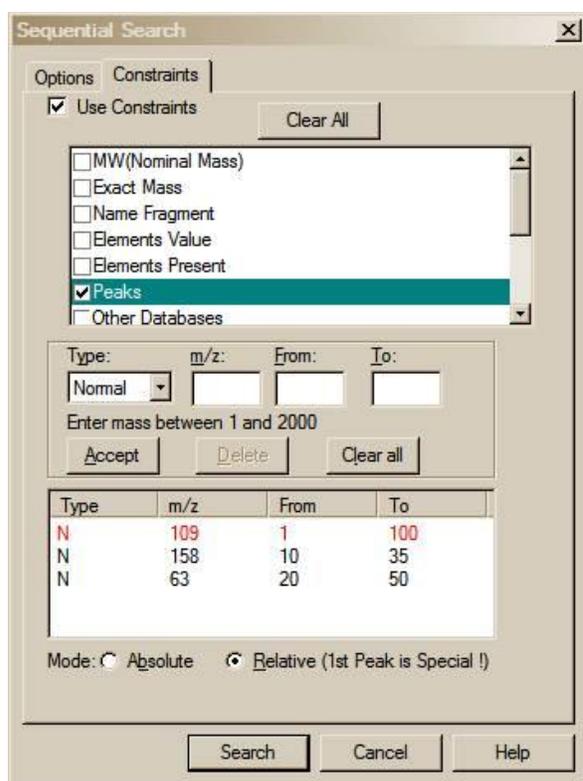


Figure 36. Sequential Search dialog box showing the use of Any Peaks in a Relative mode.

In the example shown in Figure 36, the peak at 109 is allowed to be present in a range of 1 to 100% of 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 of 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 of the databases 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 Database Search Results

When an **Identity Search** with an unknown spectrum (**User 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 database spectrum (direct match), 2) a match factor for the unknown and the database spectrum ignoring any peaks in the unknown that are not in the database 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 fairly 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** pane and on the **Title Bar** of the **Lib. Search** tab) require more explanation.

The **Probability** value for a Hit is derived assuming that the compound is represented by a spectrum in the databases 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 Database** with a set of replicate spectra (given in the **Replicates Databases**). The relative probability of each of the Hits requires only the difference values because the total probability of the compound being in the searched databases is assumed to be one. The other factor (**InLib**) is a measure of the probability of the compound actually being in the searched databases. 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 database is large (the number of "Hits vs. their Match Factors" is displayed in the **Hit Histogram** pane located just above the **Hit List**, see Figure 3).

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 Database** and searching the **Main Database** 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 a large number of 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

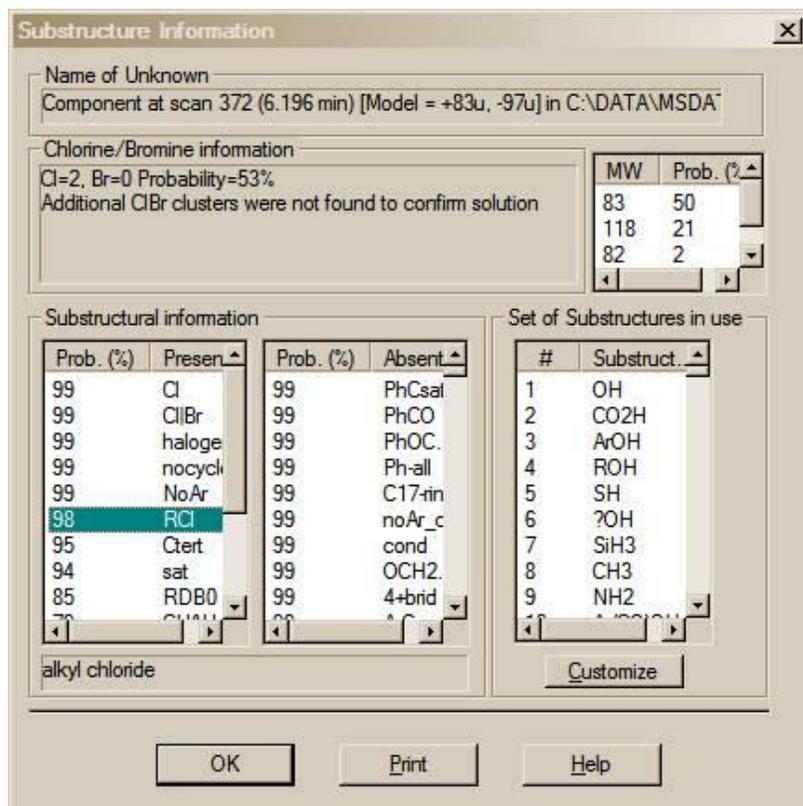


Figure 37. Substructure Information dialog box.

**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 collisionally activated dissociation (CAD) 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 37) 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 (useful for EI spectra, especially when no molecular ion peak is observed). There are two panes below this in an area labeled **Substructure information**: **Prob. Present** and **Prob. Absent**. These two panes 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** pane 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 38). This dialog box is used to create a subset of the substructures to be identified. The list of all possible substructures is displayed in

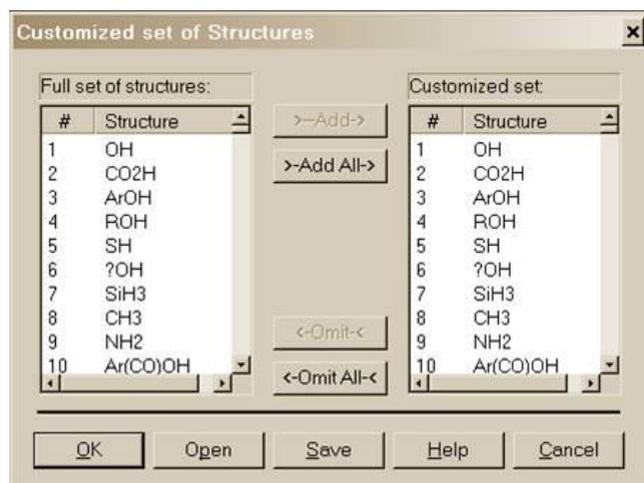


Figure 38. Customized Set of Structures dialog box.

The **Substructure Information** utility is used in the evaluation of a **Hit List** using the **NIST/EPA/NIH Mass Spectral Database** because each spectrum is indexed according to 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 (Simple, Hybrid, Neutral Loss, MS/MS in EI)**. 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 databases. Spectra similar to the searched spectrum make up the **Hit List**. There is no **Prob** column in the

---

the **Full set of structures** pane. The list of substructures included in the present file is shown in the **Customized set** pane. The first time a customized list is created, both panes will contain the same list. Substructures may be omitted from the **Customized set** pane or added from the **Full set of structures** pane. Either action occurs on the structures highlighted by Mouse action in the panes 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** pane and selecting 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).

## Example of the Use of Substructure Information in Making Determination about an Unknown Mass Spectrum obtained by using CAD

The mass spectrum in Figure 39 of a compound with the structure shown inserted in the spectrum was obtained using collisionally activated dissociation (CAD) on a protonated molecule with  $m/z$  248. An **EI in MS/MS Similarity Search** was performed against the **NIST 11 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 Database with spectra that are produced by LC/MS CAD. 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** panes 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 CAD 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.

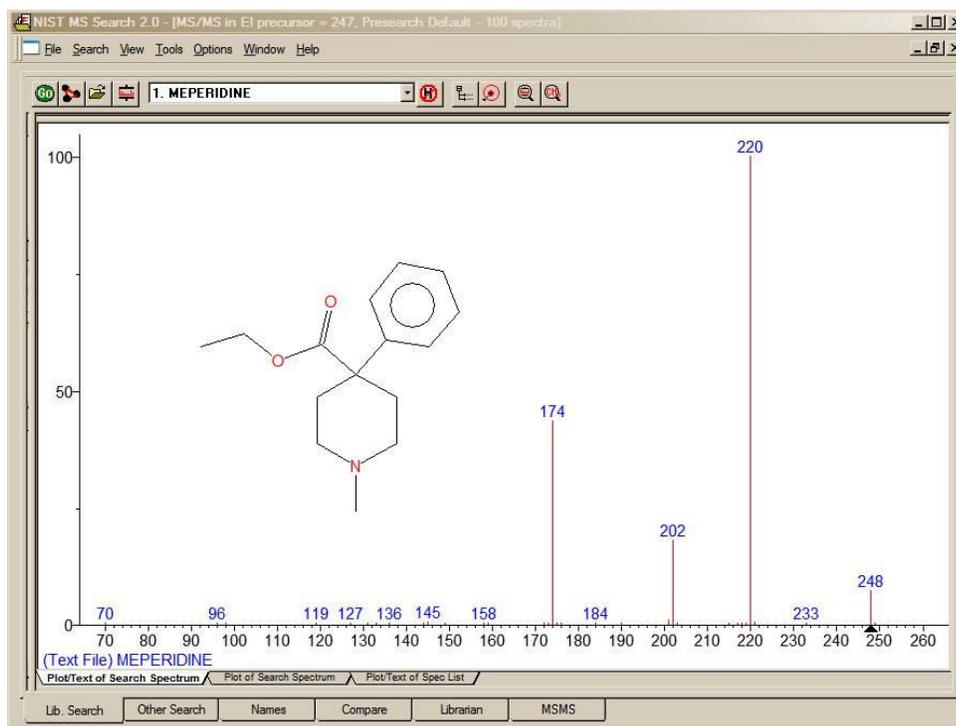


Figure 39. Mass spectrum obtained by CAD of ion with  $m/z$  248 and structure.

---

## Use with Instrument Data Systems

The use of the automation features described in the **AUTOMATIONS** section of this manual was implemented to enhance the performance of the **NIST MS Search Program** when used with proprietary mass spectral data systems.

### Location of the NIST MS Search Program

For the implementation, 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:\NIST11\MSSEARCH\
WorkDir32=C:\NIST11\MSSEARCH\
Amdis32Path=C:\NIST11\AMDIS32\
AmdisMSPath=C:\NIST11\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**.

### Implementation

If the mass spectrometry data system being used is capable of exporting spectra to the NIST format (described in the **USER DATABASES** 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:\NIST11\MSSEARCH\NISTMS$.EXE<space>/instrument
```

assuming the program is in the **C:\NIST11\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.

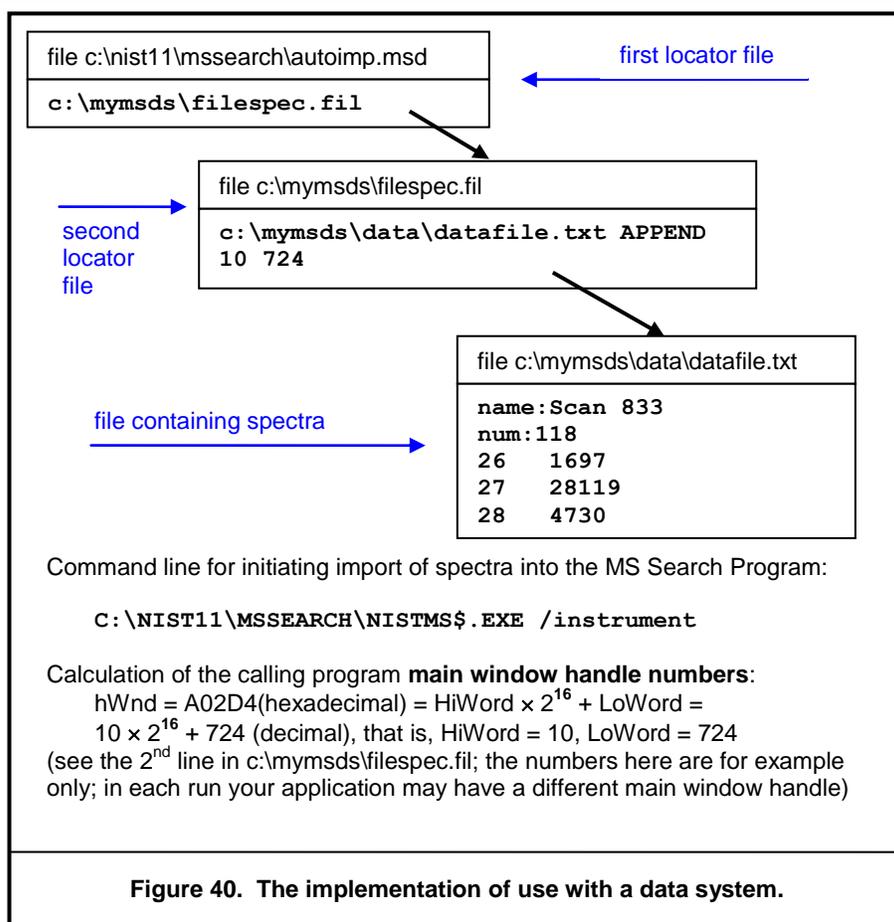
This file can be created with any text editor.

2. The second locator file is created each time spectra 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** pane 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 40 for details).

This second locator file is deleted each time the **NIST MS Search Program** is brought into focus and spectra are imported. Figure 40 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 40 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 40 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 of 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

Beginning with the first release of version 2 of the **NIST MS Search Program**, more support of structures was provided than in previous versions. There is a **Structure Search** command. If a structure is in the **Spec List** pane 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** pane 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:\NIST11\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:\NIST11\MSSEARCH\NISTMS$.EXE /MOL=<molfile>
```

assuming the **NIST MS Search Program** is in the C:\NIST11\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:\NIST02\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 Databases

One of the primary features of the **NIST MS Search Program** is the ability to do comparisons of user-generated mass spectra and create user databases to search against. The **Program** will allow up to 127 MS databases, including the databases from **NIST 11 MS Database**. The user databases may be searched in conjunction with the **NIST/EPA/NIH Mass Spectral Main** and/or **Replicates Databases**, or they may be searched separately. A maximum of 127 databases can be included for search at one time. The **Program** allows for the searching of combinations of databases in any order. The inclusion of databases and their order of search can be set from the **Libraries** or **Options** tab of the various **Search** dialog boxes (see Figure 27 and Figure 31 as examples).

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

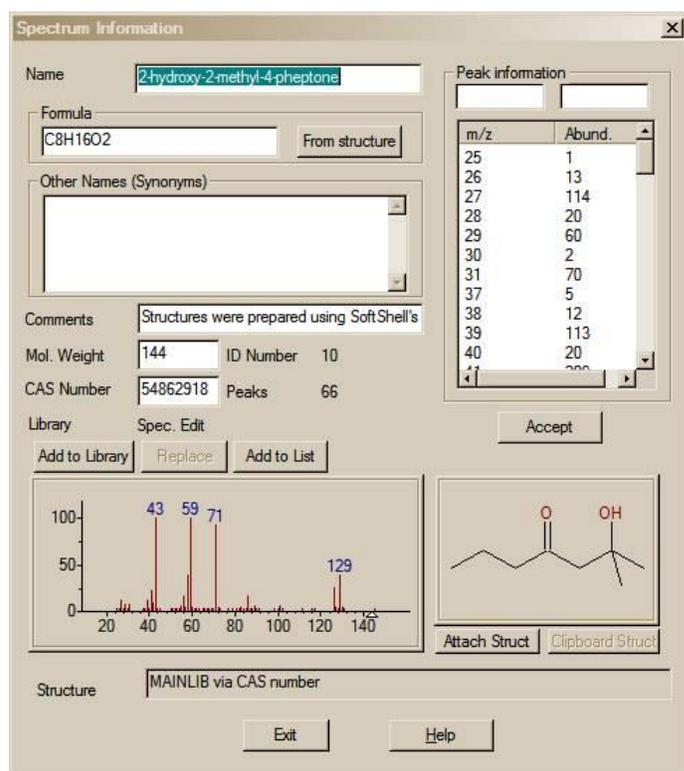


Figure 41. Spectrum Information dialog box displayed in the Librarian tab view.

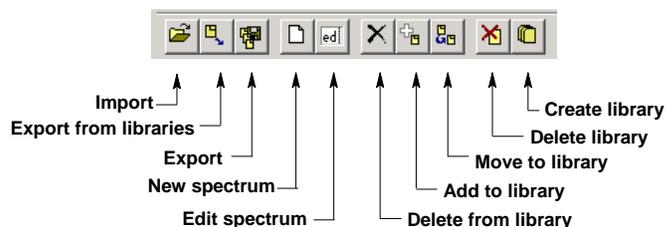


Figure 42. Librarian Tool Bar.

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. Spectra can also be entered through the **Program**. This is done by using the **Spectrum Information** dialog box (Figure 41) called with the **New spectrum** button in the **Librarian** tab **Tool Bar** (Figure 42).

User databases 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 database. Any spectrum in the **Hit List** of the **Lib. Search** tab, **Other Search** tab, or **Names** tab, a **Plot** pane, **Text Info** pane, or **Subtraction** pane in the **Compare** pane of the **Lib. Search** tab or the **Compare** tab can be copied to the **Spec List** pane by selecting **Spec List** on the **Send** to menu from the Right Mouse button menu. The size of a user database 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.

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

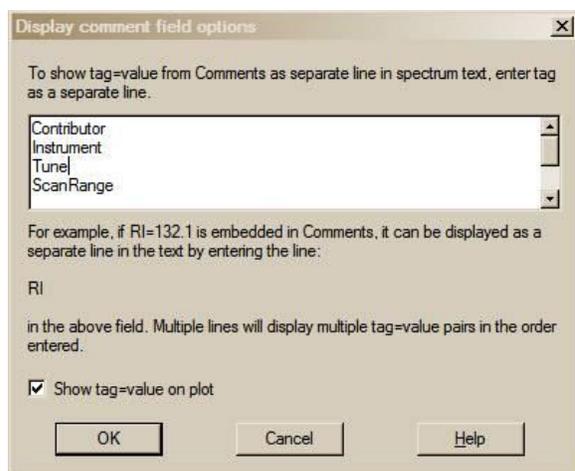


Figure 43. Display comment field options dialog box.

Introduced in v.2.0f of the **MS Search Program** distributed with **NIST 08** was the ability to interpret *Tags* in the "Comments" field of a user database spectrum. A *Tag* is the name of a Field heading that will be displayed with the *Text Information* of a user database 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 database 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 database 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 44, the information preceding each equals sign (=) will be displayed as a Field title in the **Text Info** pane of the spectrum and can optionally be displayed with the **Plot** pane 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. In order for this display to occur, the Field titles must be registered in the **Display comment field options** dialog box shown in Figure 43 (called from the **Comment field options** selection of the **View** menu). New in v.2.0g of the **MS Search Program** distributed with **NIST 11** is the feature that when the *Tags* have been entered into the **Comment field options** dialog box, they will not be displayed in the "Comments" field of the **Text Info** pane.

When a search is performed of a user database 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.

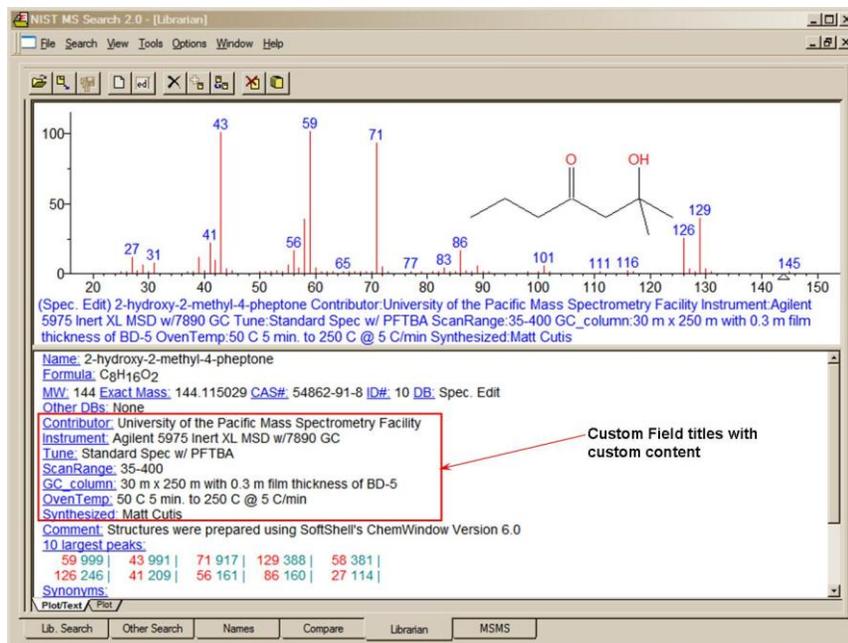


Figure 44. Example of custom Field titles with custom content added to a user library spectrum through the use of *Tags*.

---

## 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 Database**, 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 **User 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 it is present in the **NIST/EPA/NIH Main Database**.

**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) must 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-database 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.

## APPENDIX 1: Creating AUTOIMP Files Used in NIST MS Search V.2.0

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 45). 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.

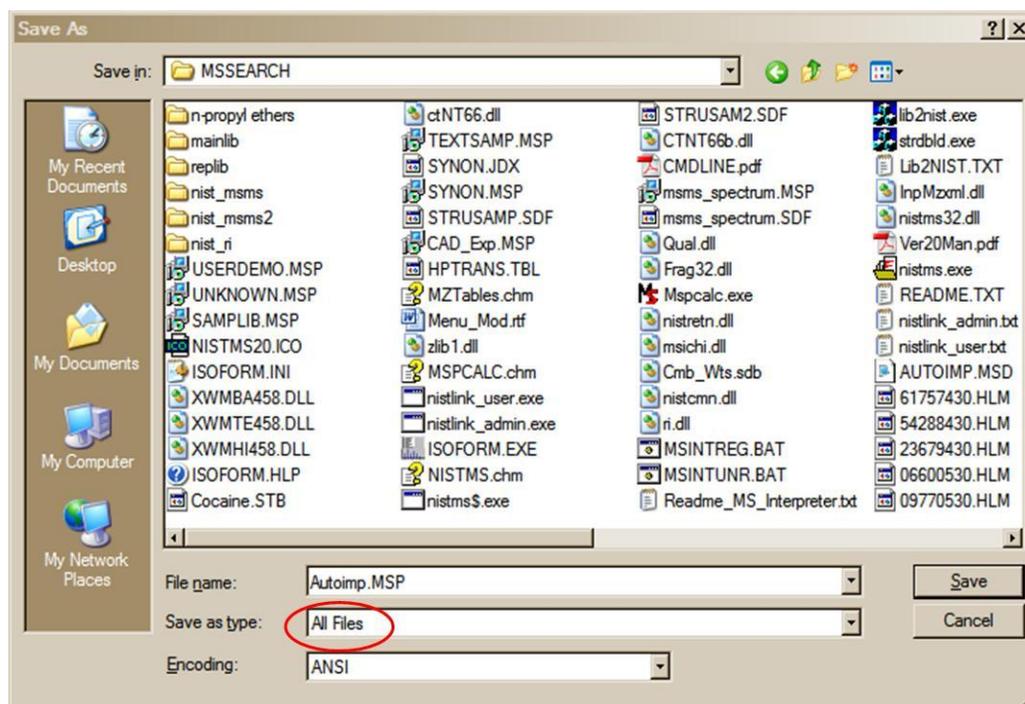


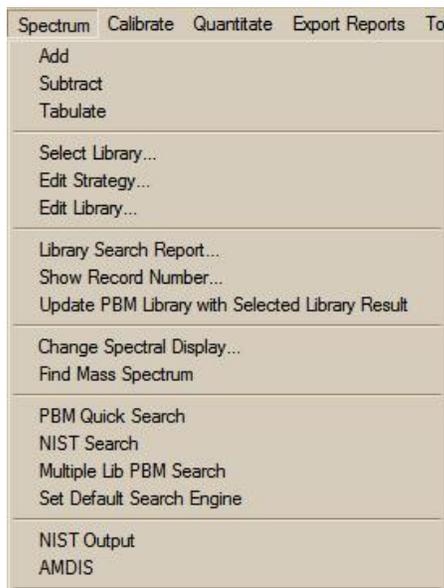
Figure 45. Save As dialog box displayed within Windows Notepad.

## 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 Database (NIST 11)** 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 26 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 11 MS Database** by selecting **ChemStation Menu Modification** from the **NIST Mass Spectral Database** program group.



**Figure 46. 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 HPCHEM\MSEXE or MSDCHEM\MSEXE 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 46):

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 databases (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 database 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 database search to be evoked when double clicking the RMB 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 databases as **NIST MS Search Program** user databases.

**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 databases that have been created using the internal utilities of the ChemStation software (including structures) and other mass spectral databases to be copied to the NIST user databases 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 group.

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 Databases (NIST 11)**, 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 group.

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 47). This dialog box consists of six options and **OK** and **Cancel** buttons. Each option 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.



**Figure 47. Result of selecting NIST Output from the Spectrum menu.**

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. In order 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 Database Format

This is a separate program called by selecting the **Lib2NIST Converter** icon on the **NIST Mass Spectral Database** program group. This program allows mass spectral databases in a variety of formats (along with structures or structures in separate files) to be selected and copied to the NIST user database format. For all 32-bit Windows (Windows NT, Windows 98/Me/2000/XP/Vista/Windows 7) 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 databases 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**.

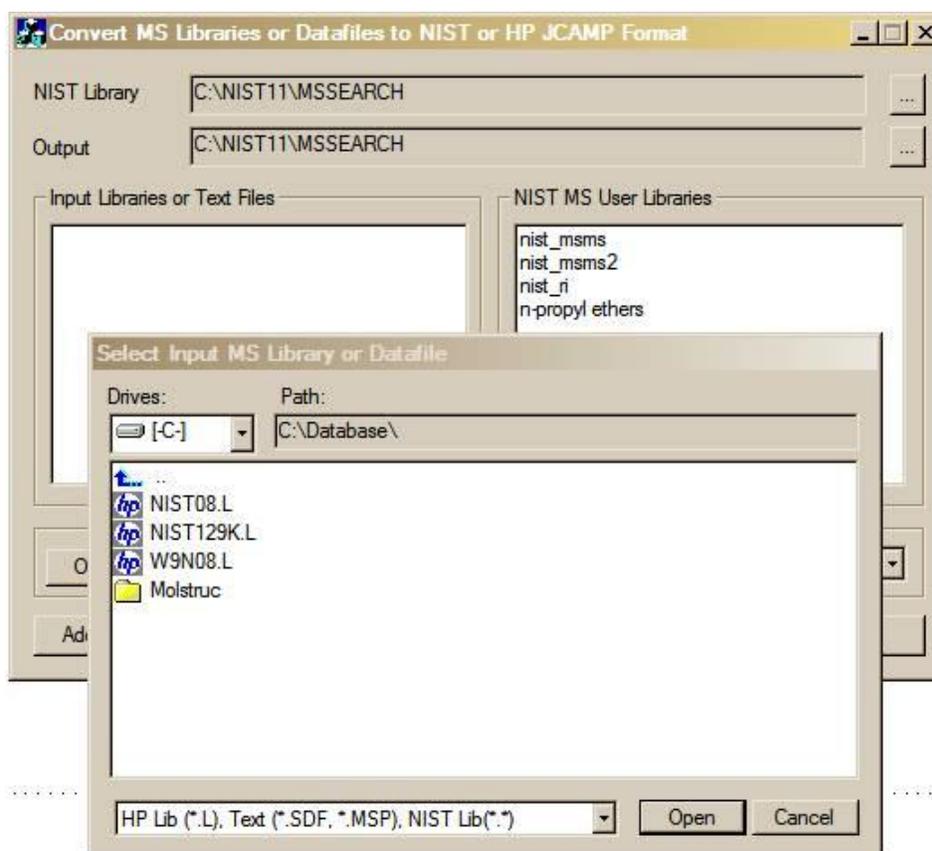


Figure 48. Opening display of Lib2NIST program. When started, the Lib2NIST converter program displays a file dialog box overlaid on the program's screen.

Select the database to be copied and then click on the **Open** button (Figure 48). The drop-down arrow will also allow for selecting from a variety of JCAMP formats. More than a single database can be selected in this dialog box by using the standard Windows multiple-file-selection techniques.

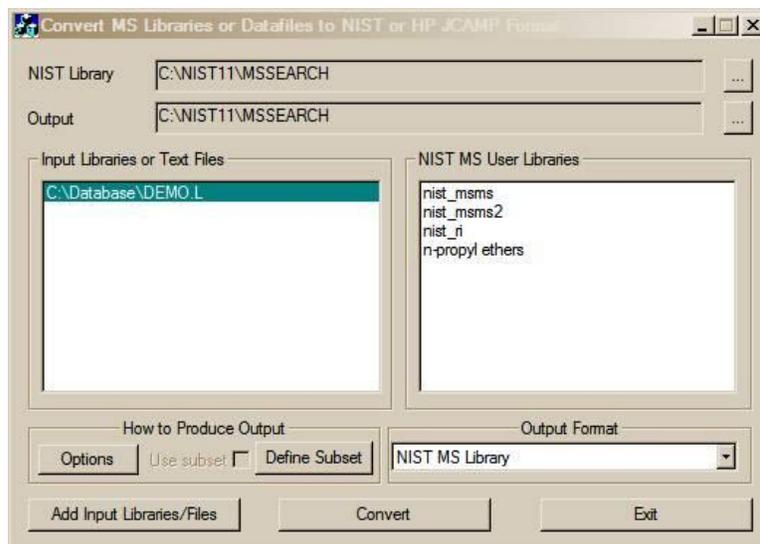


Figure 49. Lib2NIST converter program display.

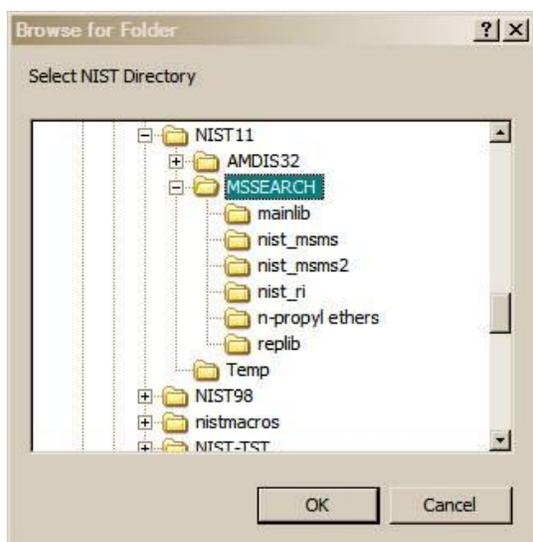


Figure 50. Conversion output selection dialog box.

The selected database(s) will now be listed in the **Input Libraries or Text Files** pane of the **Program's** display (Figure 49).

The **NIST Library** and **Output** locations can be changed by selecting the appropriate result located to the right of the two fields. This will result in the display of the **Browse for Folder** dialog box (Figure 50).

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 51).

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

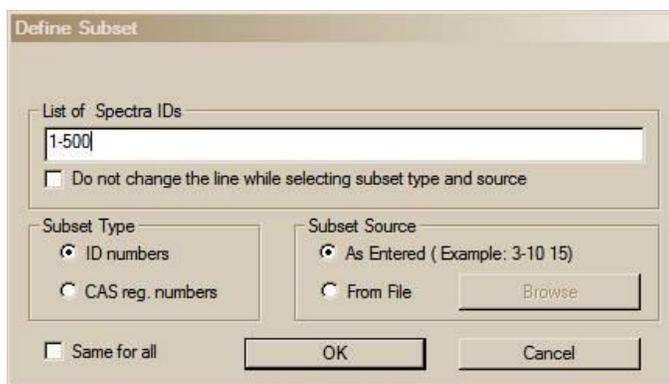


Figure 51. Define Subset dialog box.

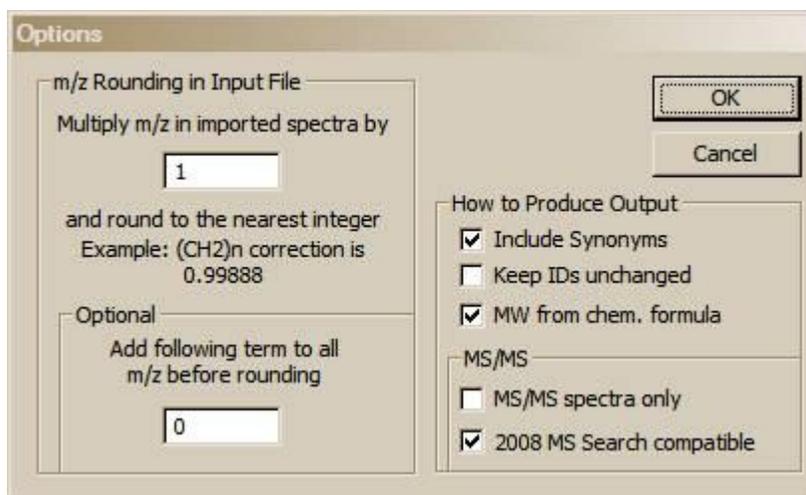


Figure 52. Options dialog box display.

The **Options** button on the **Program's** display will display the dialog box shown above in Figure 52. This allows for the entry of mass defect corrections to be applied to  $m/z$  values of imported spectra and whether or not to: 1) include synonyms from the source file, 2) retain the ID numbers (or sequence numbers) from the source database, 3) calculate molecular weights (nominal mass) from formula, and 4) restrict the output to MS/MS spectra only.

Select the databases to be converted in the **Input Libraries or Text Files** pane of the **Program's** display, make sure the Output Format is selected correctly, and then click on the **Convert** button. The selected databases will then be converted and placed in the specified output directory. Advanced features of Lib2NIST are explained in the file CMDLINE.rtf 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.

---

## 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 database 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 Database** may be optionally provided by the Thermo Fisher Corporation.

To add the **NIST/EPA/NIH Mass Spectral Databases** to the Xcalibur software, run the **NIST 11 MS Database** 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 11 in these directories. The **NIST 11 MS Database** 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:\nist11\mssearch) and the **NIST/EPA/NIH Mass Spectral Database** 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.

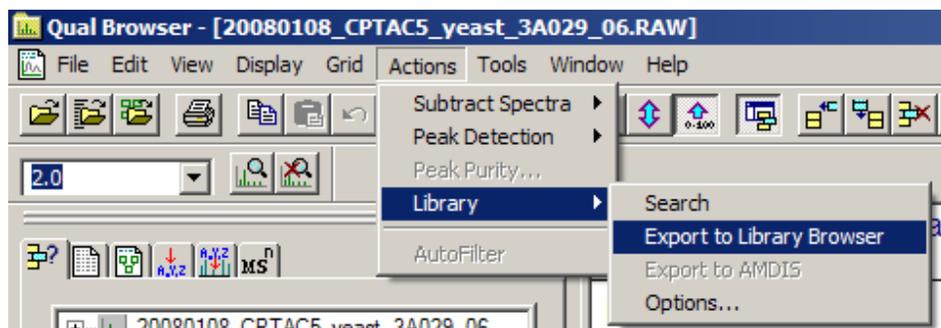


Figure 53. Exporting a spectrum to NIST MS Search.

---

## APPENDIX 5: Search Algorithms

There are two general ways that the **NIST MS Search Program** can retrieve database 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 database. 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 database).

### Screening

For the sole purpose of achieving rapid retrieval rates before actually comparing spectra, modern search algorithms first identify a subset of database 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 Database**. The best performing algorithm used “a ranked peaks in common” logic similar to that incorporated in existing data systems. This finds database 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 **Normal Identity Search**. 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 database spectra. The second most intense peak in the scaled unknown spectrum is then compared against the nine most intense peaks in the database 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 database spectra. This is the only screening procedure applied in the **Quick Identity Search**.

**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 **Quick Search**. The others were:

- 1) The fourteen largest peaks in the scaled unknown spectrum were matched against the fourteen largest peaks in the scaled database spectra.
- 2) The six largest peaks in the original (non-scaled) spectrum were matched against the six largest peaks in the original database spectra.
- 3) The five largest peaks along with the “maximum mass” peak in both the unknown and database 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** and **Neutral Loss Search**, up to five neutral loss peaks within 64  $m/z$  of the molecular ion are used in place of conventional peaks. For neutral loss peaks, abundances in database 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 database 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 database 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 databases 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 databases, 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 databases 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 &gt;= 6.8</i> <i>Diffmod &lt; 6.8</i> <i>Diffmod &lt;= 6.8</i>	<i>Diffmod=</i> followed by a number greater than 6.8 <i>Diffmod=</i> followed by a number not less than 6.8 <i>Diffmod=</i> followed by a number less than 6.8 <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	"Tag"	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	Tag="string "	exactly this <i>string</i> tagged with this Tag=
3b	Tag=""	This <i>Tag</i> has no <i>string</i> or its <i>string</i> is ""
4	Tag= <i>substring</i>	this <i>substring</i> in the <i>string</i> tagged with this Tag=
5a	Tag : <i>word</i>	exactly this <i>word</i> inside <i>string</i> tagged with this Tag=
5b	Tag : "word"	exactly this <i>word</i> inside <i>string</i> tagged with this Tag=
6	Tag= <i>number</i>	tagged string = <i>number</i> or <i>string</i> begins with <i>number</i>
7	Tag= <i>number:number</i>	tagged string <i>number</i> inside the given range
8a	Tag >= <i>number</i>	The number in the string less or equal to <i>number</i>
8b	Tag > <i>number</i>	The number in the string less than the <i>number</i>
8c	Tag <= <i>number</i>	The number in the string greater or equal to <i>number</i>
8d	Tag < <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*".

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**

	<b>Item</b>	<b>Comments of the output spectra must have</b>
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`

---

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

Text information displayed for **NIST 11 MS/MS Database** mass spectra has special tagged text fields, for example:

Instrument type: QqQ/triple quadrupole Spectrum type: ms2 Compound type: M Precursor type: [M + H] <sup>+</sup> Precursor m/z: 103.1 Collision energy: 12V Instrument: Micromass quattro Sample inlet: direct flow injection Ionization: ESI Ion mode: P Collision gas: Ar Cone voltage: 20V
---

In MS Search v.2.0g, **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 (no space) followed by the tag displayed in the spectrum text information, replacing spaces, if present, with underscores. The syntax is almost the same as the one described above for **Tags in Comment** Constraint. The search string and the special tag are case-insensitive.

Below are several examples.

To search for	Enter into the Tags in Comment Constraint
<i>precursor m/z</i> in range* 102.9–103.2	<code>\$Precursor_m/z=102.9:103.2</code>
<i>Instrument type</i> containing word QqQ	<code>\$Instrument_type:QqQ</code>
the exact <i>Instrument</i> name	<code>\$Instrument="Micromass quattro"</code>
spectra that have any Related CAS# entry	<code>\$Related_CAS#</code>
spectra that have no Precursor m/z entry	<code>^\$Precursor_m/z</code>

\* Note: Spectrum with text information "Precursor m/z: 103-105" also will be found.

### List of all acceptable special tags

\$aux	\$ion_mode	\$pressure
\$charge	\$ion_mw	\$related_cas#
\$collision_energy	\$ion_name	\$salt
\$collision_gas	\$ionization	\$salt/mix_cas#
\$compound_type	\$known_impurity	\$sample_inlet
\$cone_voltage	\$mass_range	\$special_fragmentation
\$instrument	\$maximum_intensity	\$spectrum_type
\$instrument_type	\$precursor_m/z	
\$ion_formula	\$precursor_type	

Some of these tags may be found in other **NIST 11 Databases**, for example, \$related\_cas#.

---

## ***MS/MS Instrument Type Constraint***

This feature will only work with **NIST 11 MS/MS Databases**.

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 Database does not affect the Hit List.

This constraint has 5 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 an MS/MS mass spectrum text information is interpreted in the following way:

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

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

The version of **Lib2NIST** included in **NIST 11** or a later version is needed to produce such a database because the instrument type is saved in the **NIST 11 MS/MS** mass spectrum record as a new numeric field; therefore, this constraint would not work with the **NIST 08 MS/MS Database** or even the **NIST Peptide Databases** created with the previous version of **Lib2NIST**.

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

Users of the **NIST MS Search Program** are invited to submit mass spectra for inclusion in the **NIST/EPA/NIH Mass Spectral Database**. 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 database). 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	Kovats 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.

Submissions and inquiries may be directed to Anzor I. Mikaia, National Institute of Standards and Technology, 100 Bureau Drive Stop 8380, Gaithersburg, MD 20899-8380, USA. E-mail: Anzor Mikaia, [anzor.mikaia@nist.gov](mailto:anzor.mikaia@nist.gov).

---

## 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 Databases and programs for accessing them. Comments or questions about the quality of them or other Standard Reference Databases available from the NIST Standard Reference Data Program should be made by contacting:

Joan Sauerwein  
National Institute of Standards and Technology  
Standard Reference Data Program  
100 Bureau Drive, Stop 2300  
Gaithersburg, MD 20899-2300  
e-mail: [data@nist.gov](mailto:data@nist.gov)  
Phone: 1 (301) 975-2208  
FAX#: 1 (301) 925-0416  
website <http://www.nist.gov/srd>

If there are questions or problems pertaining to the mass spectral data or use of this program or to suggest improvements, contact:

Dr. Stephen Stein  
National Institute of Standards and Technology  
100 Bureau Drive, Stop 8320  
Gaithersburg, MD 20899-8320  
e-mail: [stephen.stein@nist.gov](mailto:stephen.stein@nist.gov)  
Phone: 1 (301) 975-2505  
FAX #: 1 (301) 926-4513

or

O. David Sparkman  
5354 Thunderbird Court  
Antioch, CA 94531  
e-mail: [ods@csi.com](mailto:ods@csi.com)  
Phone: 1 (925) 754-5003  
Cell: 1 (209) 483-5740  
FAX#: 1 (772) 264-6150

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

Upgrades to the **NIST/EPA/NIH Mass Spectral Database** must be purchased.