

What's New in NIST11 (April 3, 2011)

NIST11 consists of the 2011 release* of the **NIST/EPA/NIH Electron Ionization (EI) Mass Spectral Database**, the **NIST MS/MS Database**, and the **NIST GC Methods and Retention Index Database**, along with a new version of the **NIST Mass Spectral (MS) Program** (v.2.0g). Enhanced versions of **MS Interpreter** and **AMDIS** are also included with NIST11.

The **NIST/EPA/NIH Database of Electron Ionization Mass Spectra** now contains **243,893** carefully evaluated spectra for **212,961** compounds, an increase of nearly 10% from the 2008 version. This includes **23,433** new spectra of metabolites, drugs of abuse, derivatives of common compounds and many more, all measured specifically for this library. Other major enhancements have been made to the prior version including many replacements with higher quality spectra, a thorough review of chemical names and merging of the previous salts library into the main library.

The **GC Method/Retention Index Database** now contains 349,103 GC Methods with RI values for 71,169 compounds, a 62% increase in the number of compounds in NIST08. The number of compounds with GC Methods/RI Database in the EI MS Database, (mainlib) is 38,639, an increase of 76%.

The **NIST Database of MS/MS spectra** has undergone an even greater enhancement. The new collection more than doubles the number of compounds represented. Further, most spectra have been acquired on both ion trap and qtof/triple quad instruments, thereby increasing the number of spectra by over a **factor of six** compared to the 2008 version. Spectra for the latter instrument classes have been acquired over a wide range to energies to ensure matching regardless of instrument collision settings. Also, when available, **high mass accuracy spectra** are stored. New spectra include metabolites, peptides (biologically active peptides and all di-peptides and tryptic tri-peptides), contaminants, lipids and more. The near-final version contains **9,934 ion trap spectra** for **9,138 different ions** of **4,649 compounds**, and **91,557 collision cell spectra** (qtof and tandem quad) for **6,939 different ions** of **3,774 compounds**.

There have been a number of enhancements made to the **NIST MS Search Program**; not the least of which is the addition of an **Exact Mass Search** and the ability to sort **Hit List** by the number of synonyms and other databases (non-mass spectral databases) in which the compound can be found. Features introduced in **NIST08**, like the ability to use the EI Database and the **Substructure Identification Tool** in the interpretation of CAD spectra and the ability to display isomers and derivatives as replicate spectra, have been retained and are still available. Figure 1 shows one view of the new look of the *Hit List* window and *Text Information* window in the **Lib Search Window** of MS Search v.2.0g.

The **NIST MS Search Program**, since its introduction as a Windows version (v.2.0), has evolved into a power tool for not only the matching of spectra of unknown compounds against spectra for that compound in the **EI** and **MS/MS Databases**, but also the identification of unknowns encountered in toxicology, forensic, quality control, flavor and fragrances, environmental and many other fields through their mass spectra, regardless of the type of ionization or the analyzer used to determine the abundances of ions sorted according to their *m/z* values. The **EI Database** contains much more than just the mass spectra of compounds. It contains a primary and in many cases multiple synonyms including common and trade names. It contains names of other databases where specific compounds can be found, the name of the contributor, the compound's elemental composition, its nominal and exact mass, and the compound's chemical structure.

* numbers will increase in the final version

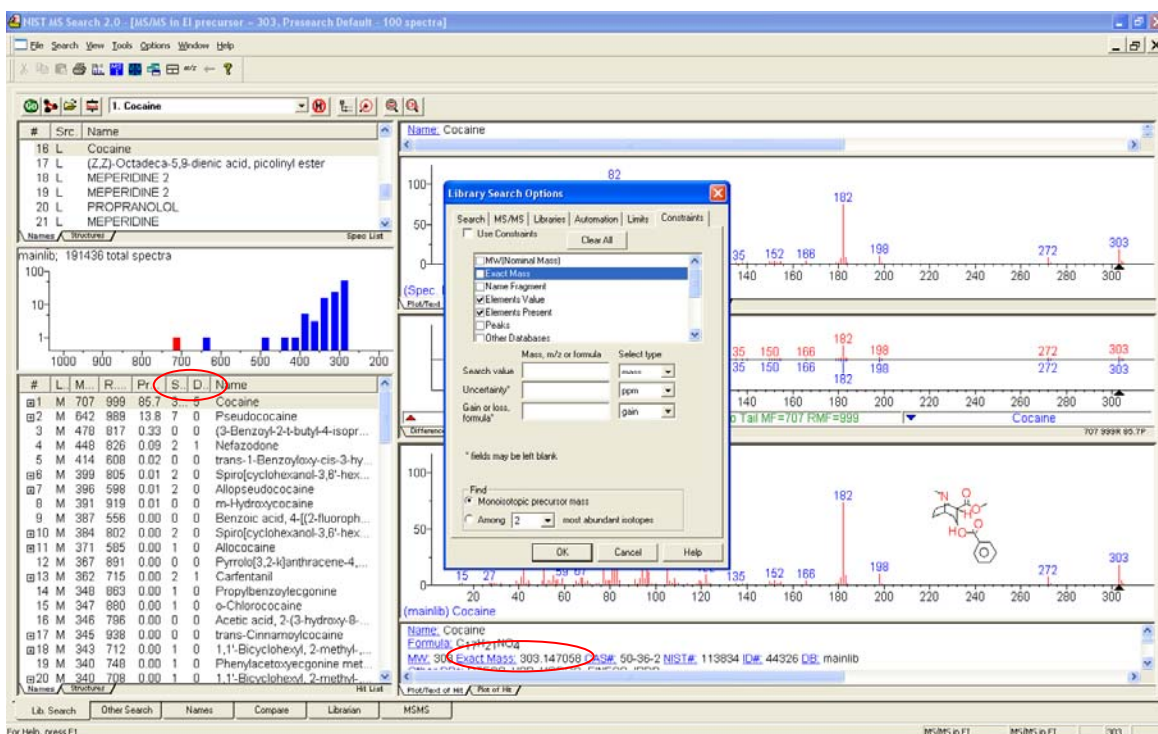


Figure 1. Library Search Window of NIST MS Search provided with NIST11 with the Exact Mass Constraints dialog box overlaid. Note the Exact Mass field in the displayed Hit List spectrum and the two new columns for the number of synonyms for the compound and the number of other databases (non-mass spectral) containing the compound.

The **NIST MS Search Program** can be an invaluable tool in GC/MS method development through the use of the **GC Method/RI Database**. This utility can be enhanced by the addition of the new **NIST GC Method/RI Search Program**, which can be used by itself in the GC laboratory. The NIST MS Search Program has evolved out of the needs of those in the NIST Mass Spectrometry Data Center that use it and the **Databases** on a daily basis. Many user-requested enhancements have been added to the **Program** which is distributed with NIST11.

The following lists a number of new features available in the **NIST MS Search Program v.2.0g**. Only by exploring, using the Help available in the Program and looking at the electronic manual that is provided will it be possible to begin to see all the features and benefits. Only through extensive utilization and practice will it be possible to see the entire picture available with the **NIST MS Search Program**.

1. The **NIST MS Search Program v.2.0g** which accompanies **NIST11** has had an **Exact Mass Search** added to the list of **Other Searches**. It takes into account the presence or absence of an electron, adducts; accepts uncertainty in ppm or millimass units, and allows searching for isotopic peaks or monoisotopic peaks. This has been enabled for both molecular ions AND product ions.
2. An **Exact Mass Constraint** has been added to a **Spectrum Search**, **Sequential Search**, and **Any Peaks Search**. This **Constraint** (Figure 1) has the same properties as the **Exact Mass Search**.

3. The **MS Interpreter** program now enables high mass accuracy for both molecular ions **AND** product ions.
4. Any mass spectral database in the **NIST MS Search Program's format** (including all user libraries, the *Wiley Registry*, and Wiley's boutique libraries like *MPW*, *DD20XX*, *FFNS*, the *Agilent DRS* libraries, etc.) can be indexed to allow use with the **Exact Mass Search** or **Exact Mass Constraint**.
5. As long as a chemical formula is associated with a mass spectrum, not only is the nominal mass (MW) displayed but the **Exact Mass** is also displayed in the spectrum's **Text Information** window.
6. Optionally displayed columns in **Hit Lists (MS Search and Other Search)** have been added for the number of **synonyms** and the number of **other databases** that a compound from the **NIST EI Database** is in; the **Hit List** may be sorted according to the contents of these columns (Figure 1).
7. Sort **Hit Lists** alphabetically (This was a feature in the *MS Search Program* in *NIST08*.)
8. Sort **Spec List** alphabetically or by number. (New to *NIST11*)
9. Toggle between **Tab** views (**Lib. Search**, **Other Search**, **Name**, **Compare**, **Librarian**, and **MSMS Windows**) using **Ctrl-Tab** or **Ctrl-Shift-Tab**.
10. The **NIST MS Search Program** will now accommodate as many 127 separate databases instead of 16, which was the limit in previous versions of the *NIST MS Search Program*.
11. The **NIST MS Search Program** will now accommodate up to 1,048,560 spectra in a database instead of 786,420 spectra allowed in the previous version.
12. **NIST11**, like **NIST08**, is fully compatible with Microsoft® *Vista™* and *Windows 7™*.
13. **Tags in Comments Field:** The concept of Tags (described below) was added to v.2.0f of the **NIST MS Search Programs**. Tags in the **Comments** field of a spectrum can be searched. With respect to the spectra in the **mainlib** and/or **replib** Databases, when using the **Sequential Search**, by typing a text string in the dialog box as shown in Figure 2, the **Contributor** field will be searched. All spectra contributed by a single source can be identified in this way. This feature is continued v.2.0g and even further enhanced.
14. When the **Mouse** pointer is in a **Text Information** window, clicking the **Right Mouse** button will display a popup menu, there is a **Find** option that allows the window to be searched for a specific text string. This is particularly useful with **GC Method/RI** records that are associated with spectra in the **NIST/EPA/NIH Database**.
15. One or more items from any **Hit List** may be copied into the **Windows Clipboard** as tab-delimited **Unformatted Unicode** text and pasted into **Excel™**. The copying may be done from the **Right-Mouse-button Menu** or by pressing **Ctrl-K** (possibly subject to change) with the spectrum (spectra) highlighted. To select all the spectra in a **Hit List**, press **Ctrl-A** (a feature from *NIST08*). Note: Plain **ASCII** text, which may be obtained using **Paste Special/Unformatted** text, would not have proper **Greek** letters.

16. Tags (Field titles) in User Library & Spec List Spectra

A *Tag* is the name of a Field heading that will be displayed with the *Text Information* of a User Library or Spec List spectrum and can optionally appear in the line below the *Plot* of the spectrum. Text is designated as Tag in the **Comments** field of a User Library or Spec List spectrum with the *Spectrum Edit* function of the Librarian Window by entering the Field title (a line of characters without spaces) followed by an =. Following the = is the contents of Field. This can be a line of characters without spaces, a number or a text string with open and close quotation marks (""). For example, the following may be entered in the **Comments** field of the **Spectrum Information** dialog box of a Spec List or a User Library spectrum:

Structures were provided by O. David Sparkman
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 Cutis"

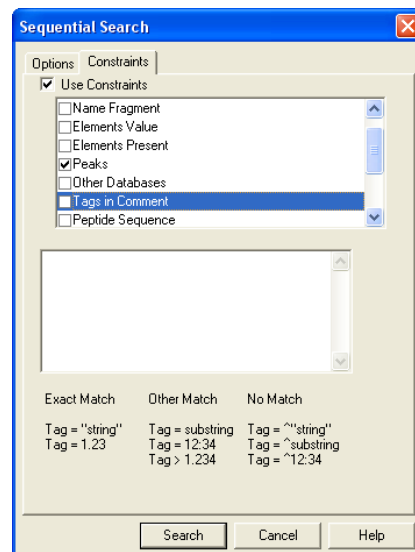


Figure 2

As shown in Figure 5, the information preceding each equal sign (=) will be displayed as a Field title in the Text display of the spectrum and can optional be displayed with the Plot 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 4 (**View/Comment fields options**). 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 Information window. Figure 5 shows a spectrum (Plot and Text) with added Field headers.

When a search is done of a User library that contains *Tags* (Field titles), it is possible to Constrain that search based on the *Tag* (Field title) and the partial or whole contents of the *Tag* (Field title). More specific details on the use of Tags as Constraints are in MS Search Help and Manual.

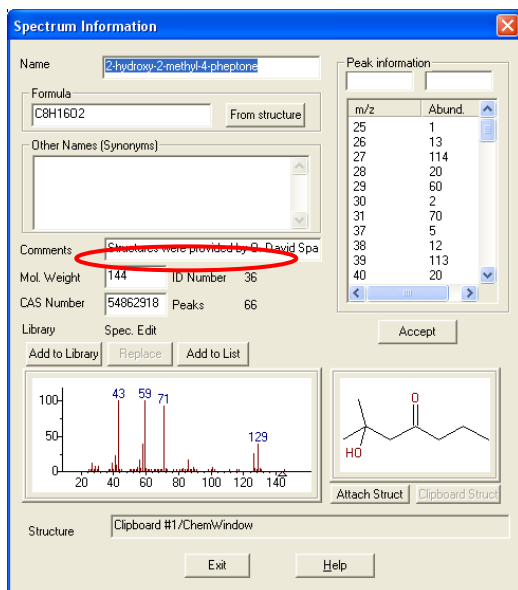


Figure 3

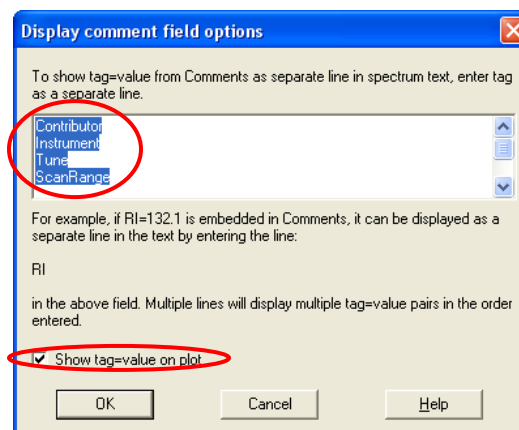


Figure 4

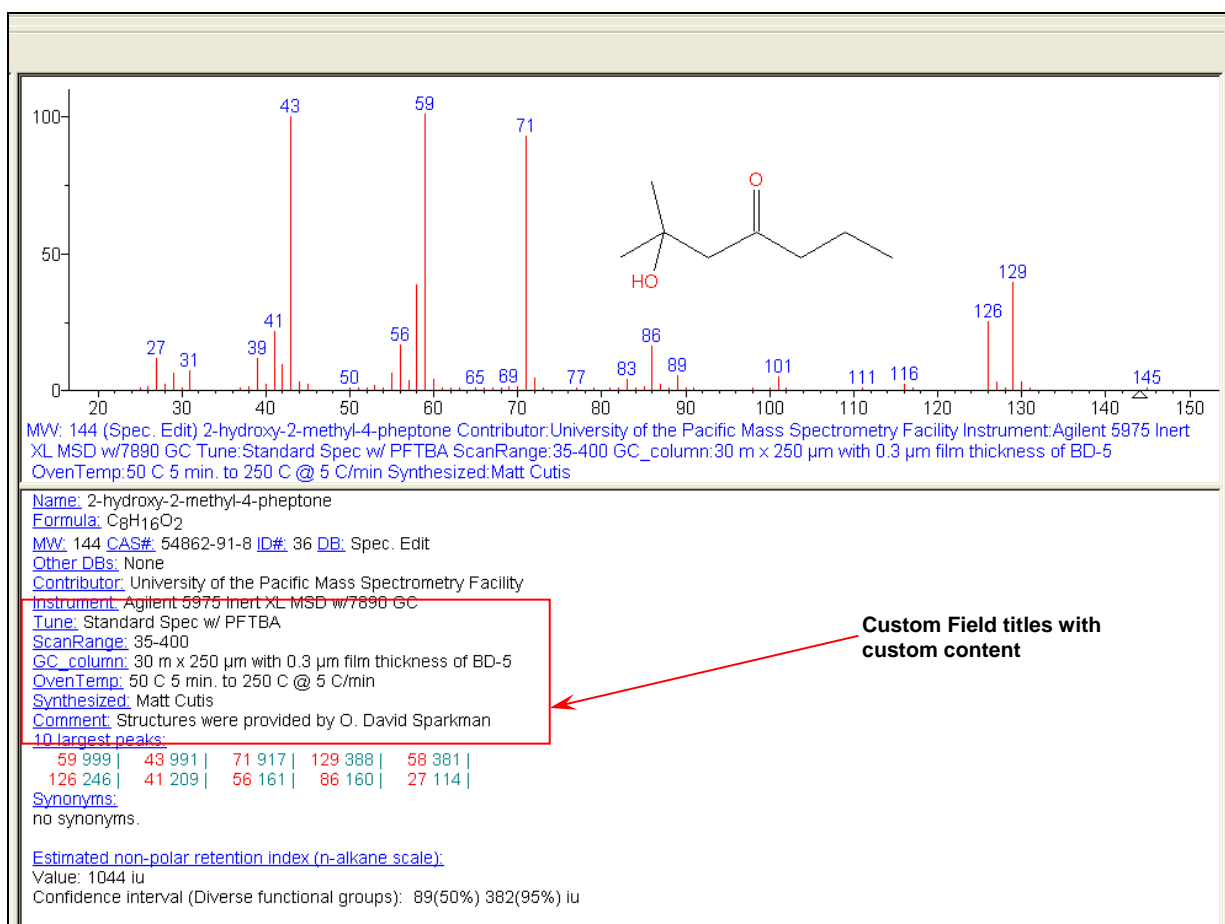


Figure 5

17. **Display of Spectra of Derivatives as Replicate Spectra:** One of the major new features with NIST MS Search (v.2.0f) in NIST08 was the ability to display spectra of derivatives as replicate spectra. If a spectrum of 1-octanol is in a Hit List and the **mainlib** or **mainlib** and **replib** databases have spectra of trimethyl silyl, pentamethyldisilyl, and/or acetate derivatives of 1-octanol, these would be listed as replicate spectra. When the replicates are displayed, actual replicates of compounds found in the **replib** database would be designated with an **R**. The spectra of the derivatives would be designated with either an **m** (in the **mainlib** database) or an **r** (in the **replib** database). This display feature is defaulted in MS Search 2.0f. To turn this feature off or modify it, use the **Replicates Display Options** dialog box called by **Options/Replicates**. This feature which first appeared with NIST08 has been enhanced for NIST11.
18. The *MS/MS in EI* option in the *Spectrum Similarity Search* is used with spectra of unknowns obtained by fragmentation process, EI or Collision Activated Decomposition (MS/MS spectra) for the determination of the probability of the *Presence* and *Absence* of substructures using the NIST companion **Substructure Identification Tool**. The *MS/MS in EI* option accesses neutral losses in the sample spectrum. A search against the neutral losses of the spectra in the **NIST EI Database** is then carried out. A **Hit List** containing EI spectra which have exhibited the same or similar neutral losses from the molecular ion of compounds in the EI Database is generated.

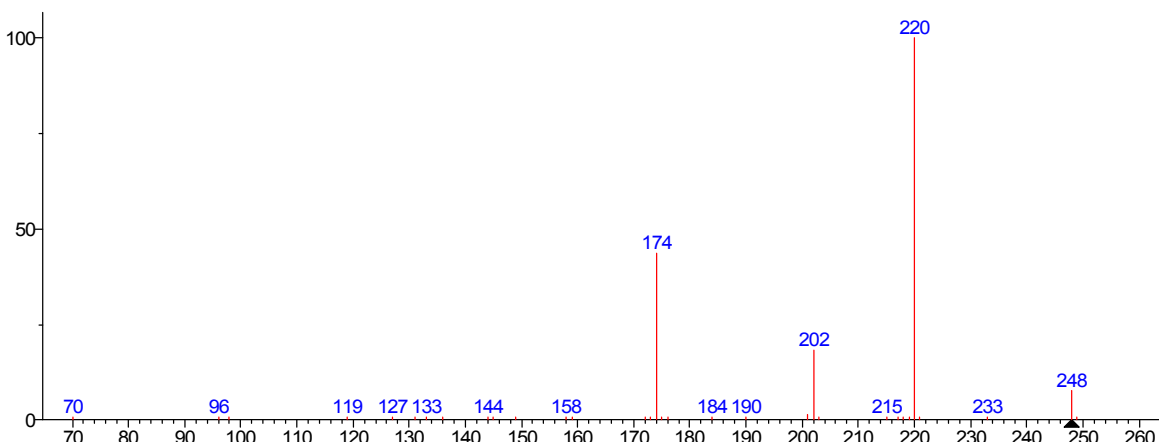


Figure 6. A spectrum obtained by MS/MS using the protonated molecule at m/z 248 obtained by electrospray as the precursor ion.

Each spectrum in the EI Database is indexed according to substructures of the compound that generated the spectrum. Substructures are features such as a functional group, heteroatoms, the number of rings-plus-double bonds (R+dB), etc. Figure 6 is a mass spectrum acquired using MS/MS in a 3D Quadrupole Ion Trap of a precursor ion with m/z 248 formed with positive-ion electrospray. This spectrum was searched using the NIST MS/MS in EI option *Spectrum Similarity Search* in the **MS Search Program** and the **NIST11 EI Mass Spectral Database**. The resulting Hit List (not of spectra of similar compounds but of EI spectra that exhibited the same natural losses as the CAD spectrum) was then analyzed using the **Substructure Identification Tool**. A portion of the Results are shown in Figure 7.

The **Substructure Identification Tool** indicated that there were probabilities $\geq 90\%$ that the unknown contained one or more atoms of oxygen, had a R+dB count greater than five, that it contained both a methyl and a methylene group and that there was a substructure with both connected to one another, and a carbonyl group. There were probabilities $>85\%$ that the unknown contained one or more atoms of nitrogen, that there was an aromatic ring and a heterocyclic ring with nitrogen, a carbon oxygen bond, and many other substructures.

The probability of the absences of substructures showed a $>90\%$ probability that there was not a phenyl carbonyl or phenyl

methoxy, a dialkyl ether, an oxygen in a ring, benzyl moiety, an alkyl-aryl ether, or an aryl nitro group.

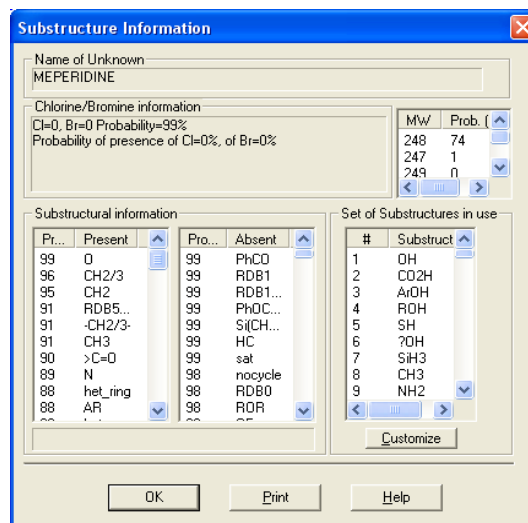
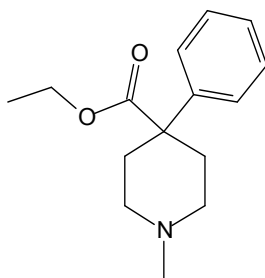


Figure 7

Using these reported presences and absences it was not too difficult to establish a probable structure as shown to the left. Not only can the Substructure Identification tool be used in the evaluation of CAD (MS/MS Spectra), it can also be used in the identification of unknowns using spectra when there is no spectrum for the unknown in a database.

19. When performing **Other Searches** such as **Molecular Weight**, **Sequential**, and **Formula Searches** it may be desirable to change or include Constraints to limit the number of hits. This may need to be done for several iterations. To facilitate this, *now* when one of the **Searches** is requested the Search dialog box opens with Constraints tab display rather than the options tab (See Figures 8 and 9).

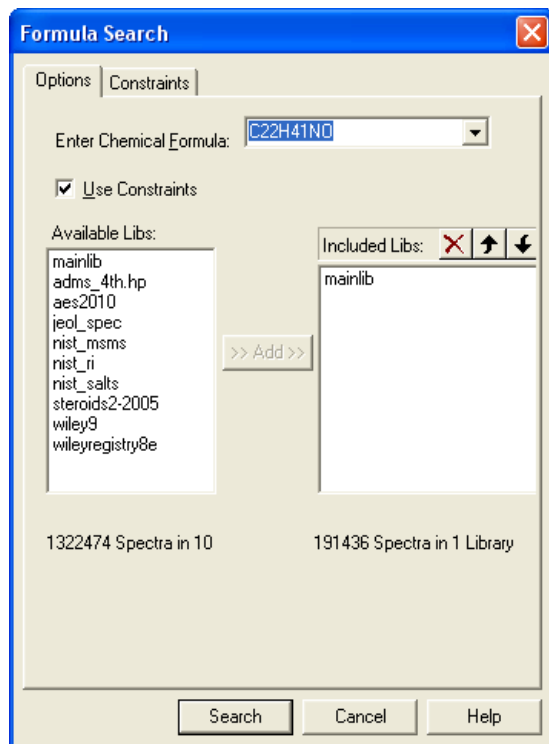


Figure 8. Search Dialog Box Displayed on Opening in NIST08

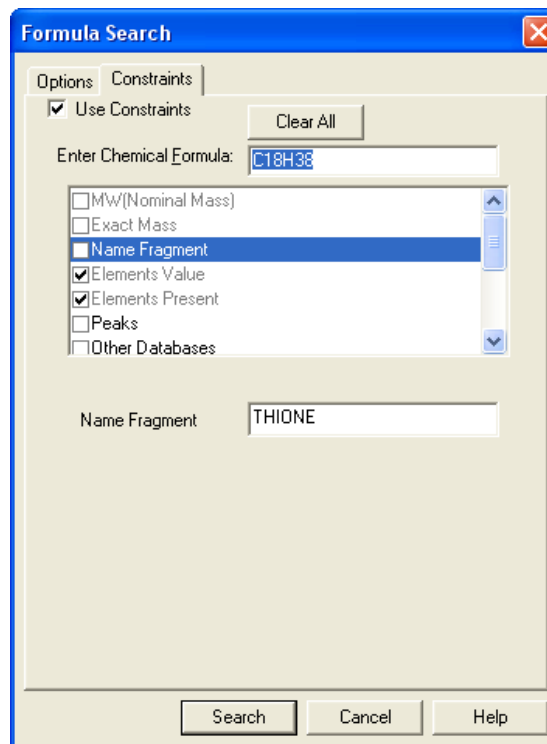


Figure 9. Search Dialog Box Displayed on Opening in NIST11

20. A possibility to turn off Homolog hits in a *Structure Similarity Search* has been added. This is a new feature in the Library Search Options dialog box
21. Spectra in the MSP text format can be imported from the Windows Clipboard into the Spec List via a Right Mouse button Menu option.
22. In the new mainlib-11 a relatively small number of compounds have *more than one CAS registry number*. These compounds may be found by the usual CAS r.n. search; a new file registry2.in6 in the NIST11 Main and Replicate Library-folders is required. If the registry2.in6 file is not present, then a regular CAS search is performed thus providing backward compatibility. The current version does not display these extra CAS r.n.'s; although, they are present as invisible synonyms.

23. The NIST MS Search Program can import centroided spectra from **mzXML** and **mzData** MS and MS/MS files.
24. An *alternative peak matching method* has been added to improve the reliability of the score when searching noisy MS/MS spectra.
25. The **NIST MS Search Program** is fully compatible with most recent **NIST Peptide MS/MS Libraries**.
26. A chemical structure in the **MOL file** format can be inserted from the Windows clipboard into the Spec List in Librarian or Lib Search Window via a right-mouse-button menu command allowing for a *Structure Similarity Search*.
27. A chemical structure in the MOL file format located on the Windows Clipboard can be attached to the plot in *Search Spectrum* window displayed in the *Lib. Search* Window. The result may be subsequently sent to *Spec. List* and added to a user library (this undocumented feature was in the version of the **MS Search Program** distributed with NIST08 software).