

Agilent MassHunter Molecular Structure Correlator (MSC) Software

Quick Start Guide

What is MSC?	2
How can you use MSC?	3
Installation	4
Getting Started	5
To start the MassHunter MSC	5
Step 1. Load fragment spectrum	7
Step 2. Enter compound chemistry and data quality information	8
Step 3. Review molecular formulas for the compound	11
Step 4. Select a formula for fragment structure correlation analysis	12
Step 5. Define structure candidates to correlate to the selected formula and MS/MS spectrum	12
Step 6. Review the candidate structures	16
Step 7. Select a structure to view its substructure assignment	16
Step 8. Review the possible substructures for each fragment peak	17
Step 9. (Optional) Select another formula	17
Familiarization Exercises	18
Exercise 1. Confirm a proposed structure	18
Exercise 2. Identify true unknowns	19
Exercise 3. Correlate an MS/MS spectrum input via a .ms2 text file	21
Exercise 4. Select an MS/MS spectrum for correlation from a CEF file which contains multiple MS/MS spectra	21
Reference	22

This software is provided “as is” as a supplemental tool.



Agilent Technologies

What is MSC?

The MassHunter MSC (Molecular Structure Correlator) program correlates accurate mass MS/MS fragment ions for a compound of interest with one or more proposed molecular structures for that compound. MSC accomplishes this by trying to explain each observed fragment ion into the proposed structure using a “systematic bond-breaking” approach as described by Hill and Mortishire-Smith (see “Basis for MSC” on page 22).

The input for MSC is an accurate mass MS/MS fragment spectrum, one or multiple candidate formulas for the compound of interest, and one or multiple candidate molecular structures.

The user can either input a molecular formula manually, or have MSC select the *n* most probable molecular formulas it calculates for a compound using the accurate mass MS and MS/MS information.

The MSC then uses the selected formula or formulas, retrieves one or multiple possible structures from a .mol file, an .sdf file, a MassHunter compound database (PCD, PCDL), PubChem (web) or ChemSpider (web) and scores how well each candidate structure correlates with the MS/MS spectrum. The overall correlation score gets calculated from individual scores for each fragment ion signal. For each fragment ion one or multiple substructure candidates may be suggested and a “penalty” assigned based on how many and what bonds need to be broken to generate that substructure. Breaking two bonds, a double bond or even an aromatic ring carries a higher penalty (requires more energy and therefore is less likely) than just breaking one single bond. Two other factors impacting the overall correlation score are the mass accuracy of the observed fragment ions and the overall percentage of fragment ion intensity that can be plausibly explained with substructures.

MSC limits the maximum correlation score if only 1, 2 or 3 fragment ions are present in the MS/MS spectrum. MSC now also calculates a combined score from the MFG Parent score (MS level) and the overall MSC score (MS/MS level), because it might return candidate structures from multiple candidate formulas.

How can you use MSC?

Confirm a proposed structure for a compound

MSC can be used to confirm a proposed structure for a compound with higher confidence by correlating, as much as possible, fragment ions in the MS/MS spectrum with plausible substructures. The correlation score gives an indication how likely the MS/MS spectrum supports the identity of the proposed structure. MSC can be used to perform structural confirmation in applications such as drug metabolite ID, impurity analysis, metabolomics, screening in food safety, forensics/toxicology or environmental samples.

Aid in the identification of true unknowns

MSC can *aid* in the identification of true unknowns by running the correlation of the unknown MS/MS spectrum against multiple candidate structures and calculating correlation scores for each of them. MSC retrieves all possible structures for the *n* most likely formulas from a local compound database (MassHunter PCD or PCDL) or a web compound database (PubChem or ChemSpider). The candidates with higher scores are more likely to be the correct structure.

MSC can help to identify the most probable molecular formulas by using both MS and MS/MS fragment and neutral loss information in the molecular formula generation.

MSC cannot unambiguously identify a true unknown compound via its accurate mass MS/MS spectrum, because while there might be only ten possible candidate structures, there could also be over one hundred or even one thousand possible structures in a large compound database for the *n* possible candidate formulas. The difference between the correlation scores might not be discriminatory enough to provide an unambiguous decision between candidate structures. It can, however, be used to narrow down the possibilities, give an idea about the substructures that might exist in the unknown molecule or to suggest the class of molecules the unknown compound might belong.

Installation

- 1** Run **MSC.msi** to install the MSC software.
The files are installed in **C:\Program Files\Agilent\MassHunter\Workstation\MSC**.
- 2** Copy example data (.cef files) from the **Data** folder on the installation CD to a location of choice on your hard drive.
The default location is **C:\Program Files\Agilent\MassHunter\Workstation\MSC\Data**.
- 3** If you have not installed PCDL Manager B.07.00 (or higher) or MassHunter Qualitative Analysis B.07.00 (or higher), install the SqlServer Compact Edition 3.51 (SSCE) and both the 32-bit version (SSCERuntime-ENU-x86.msi) and the 64-bit version (SSCERuntime-ENU-x64.msi).

Getting Started

To start the MassHunter MSC

Start the MassHunter MSC in one of two ways:

- Double-click the desktop icon, *or*
- Click **Start > All Programs > Agilent > Molecular Structure Correlator > Molecular Structure Correlator**.



The MSC application has two main panes.

- The Spectrum List Pane (left) displays the list of spectra that have been loaded. The list shows for each spectrum the RT, m/z , name, polarity, C.E., identifier, formula, dM (ppm), score, and StdInChI Key. To view the MSG results, right-click the front section of a row for a spectrum. The results appear in a separate window.
- The Structural Information pane (right) is divided in two sections. The left section shows the candidate structures of the parent compound (see “[Step 6. Review the candidate structures](#)” on page 16). The right section is used to review the fragment ion table and the substructure assignments (see “[Step 8. Review the possible substructures for each fragment peak](#)” on page 17).

Getting Started

To start the MassHunter MSC





The screenshot displays the Agilent MassHunter Molecular Structure Correlator (MSC) software interface. The interface is divided into several panels:

- 1 Spectrum Loaded:** Shows a table of mass spectrometry data. A red box highlights the "Spectrum List" table.
- 35 structures found for Proadifen:** Shows chemical structures and their scores. A red box highlights the text "candidate structures of the parent compound".
- Structure #1 - eluoidated: 60.0% ions, 96.4% Weight:** Shows a table of mass spectrometry data and chemical structures. A red box highlights the text "substructure assignments".


Figure 1 User Interface of Molecular Structure Correlator

Toolbar Icons

At the top are four toolbar icons:

- Click  to create a project.
- Click  to open a project.
- Click  to save the project.
- Click  to run the MSC.

Step 1. Load fragment spectrum

- 1 Click  in the main toolbar, or click **Project > Create**.
- 2 In the **Create Project** dialog box, type a name for the project.
- 3 Click **Add files**.
- 4 Specify which kind of file to open (*.cef or *.ms2).

The .cef files are created by the MassHunter Qualitative Analysis program. The format for the .ms2 file is specified in “[Structure of .ms2 file](#)” on page 22.

If the .cef file contains more than one fragment spectrum, MSC will automatically execute the structure correlation for each spectrum sequentially. You can then click the area in front of each spectrum row to review the results.

- 5 In the **Create Project** dialog box, click **OK**.

Step 2. Enter compound chemistry and data quality information

- 1 Click **Settings > Data Loading** to open the **Data Loading** dialog box.

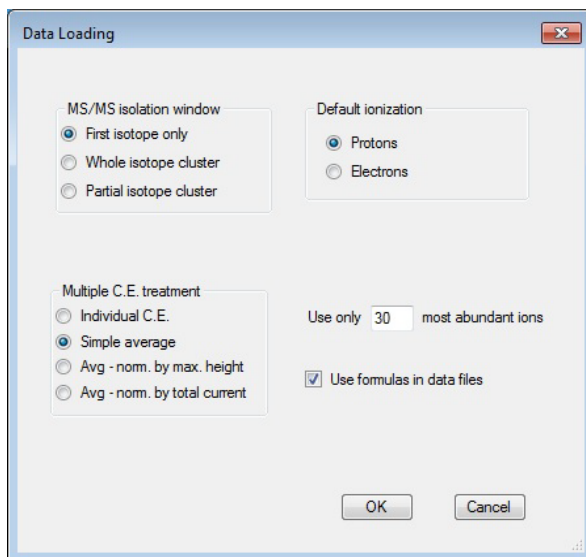


Figure 2 Data Loading spectral processing parameters

- 2 Change the options in the **Data Loading** dialog box as needed, and then click **OK**. Refer to online Help for more information.

The options for **MS/MS Isolation Window** refer to how data was acquired.

For singly charged compounds, select **First Isotope Only**. Only the first isotope is isolated, so the fragments in the MS/MS spectrum also show the first isotope only.

To interrogate a spectrum that was acquired with a wide isolation window, select **Whole Isotope Cluster**. MSC recognizes fragment ion clusters and deisotopes them.

Whole Isotope Cluster is also used for EI-MS TOF spectra. No isolation is run for EI-MS TOF spectra, so fragment ions show the entire isotope cluster and need to be deisotoped.

You can also limit the correlation to the n most abundant fragment ions (after deisotoping, default is 30).

- 3 To change the Molecular Formula Generation (MFG) parameters, click **Settings > MFG Parameters** to open the **MFG Parameters** dialog box.

The parameters in the **MFG Parameters** dialog box are similar to those used in MassHunter Qualitative Analysis for Molecular Formula Generation (MFG).

To change the allowed elements in the elemental composition, click **add/remove**.

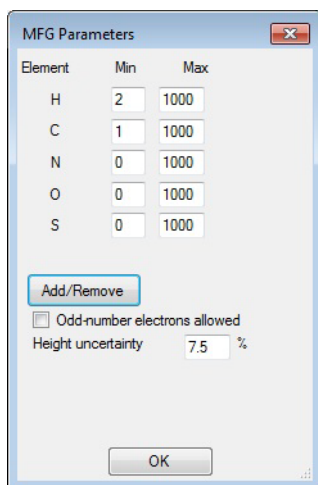
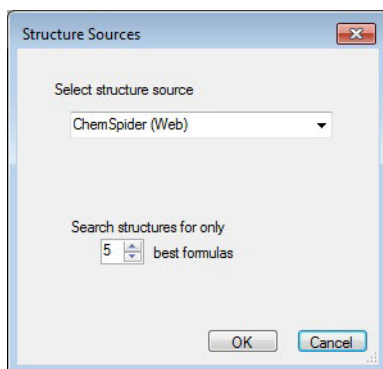


Figure 3 Setup MFG Parameters

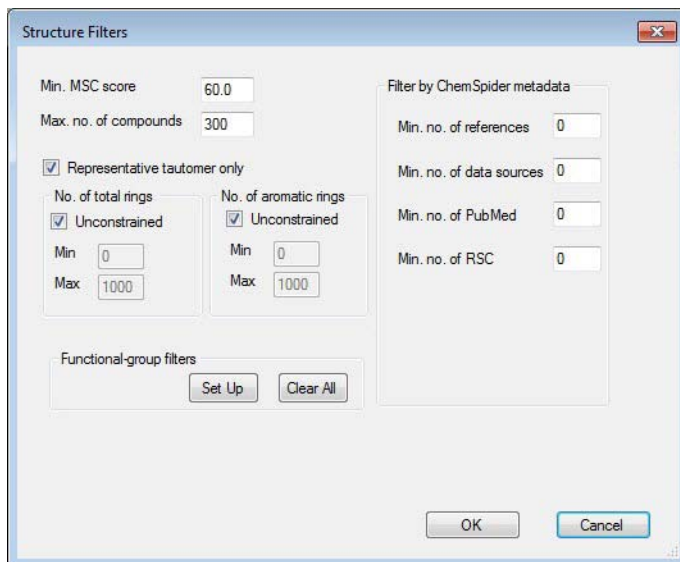
- 4 Click **Settings > Structure Sources** to change the structure sources.



Getting Started

Step 2. Enter compound chemistry and data quality information

- 5 Click **Settings > Structures Filters** to set the search parameters.



The screenshot shows the "Structure Filters" dialog box with the following settings:

- Min. MSC score: 60.0
- Max. no. of compounds: 300
- Representative tautomer only
- No. of total rings:
 - Unconstrained
 - Min: 0
 - Max: 1000
- No. of aromatic rings:
 - Unconstrained
 - Min: 0
 - Max: 1000
- Functional-group filters: Set Up, Clear All
- Filter by ChemSpider metadata:
 - Min. no. of references: 0
 - Min. no. of data sources: 0
 - Min. no. of PubMed: 0
 - Min. no. of RSC: 0
- Buttons: OK, Cancel

- 6 To process the data, click .

Step 3. Review molecular formulas for the compound

The upper pane in the **MFG Result** dialog box lists the candidate formulas that MSC calculated for the compound.

- To enter a different formula for a compound, right-click in the first cell of a row in the Spectrum List table for the spectrum/compound of interest, then select **Impose formula**, which opens the **Input Formula** dialog box. Type in the desired molecular formula, then click **OK**.

Added formulas appear in red in the table. They must equal the neutral mass of the precursor ion to be accepted. At this point, MSC does not calculate formulas for the fragments because MFG, which calculates fragment formulas, has not been rerun.

Note that when multiple plausible formulas can be generated for at least one fragment ion, a common color coding helps to recognize the fragment ion.

Refer to online Help for a description of each column.

- To open the MFG window to display molecular formulas, right-click the first cell in the row for the compound of interest in the **Spectrum List** table and select **View MFG Result**.

MFG Result

M = 270.0246; 6 formula candidates from MFG

ID	Formula	Isomers	Taut. Grps	dM(ppm)	IdM%
1	C9H10N4O2S2	74	66	-0.4	
2	C8H14O6S2	3	3	-5.3	
3	C8H6N4O7	6	6	-3.6	
4	C12H6N4O2S	11	11	-12.9	
5	C9H18O54	2	2	-2.1	
6	C13H6N2O5	18	18	11.3	

Fragment formulas for C9H10N4O2S2

m/z	intensity	nom. intens.	formula	dM(ppm)
92.0493	550.68	10.19	C6H6N	
92.0493	550.68	10.19	C3H10NS	
92.0493	550.68	10.19	CH6N3O2	
108.0440	1022.86	18.93	C6H6NO	
108.0440	1022.86	18.93	C4H4N4	
108.0440	1022.86	18.93	CH8N4S	
108.0440	1022.86	18.93	C3H10NOS	

Close

Figure 4 MFG Result

Step 4. Select a formula for fragment structure correlation analysis

- Click the row header in the Precursor formula to select the formula of interest.

The correlation results are cleared, and the newly selected formula is displayed in the header across the center and right panes.

Step 5. Define structure candidates to correlate to the selected formula and MS/MS spectrum

You can select one or multiple molecular structures to correlate to the accurate mass MS/MS spectrum, depending on your application.

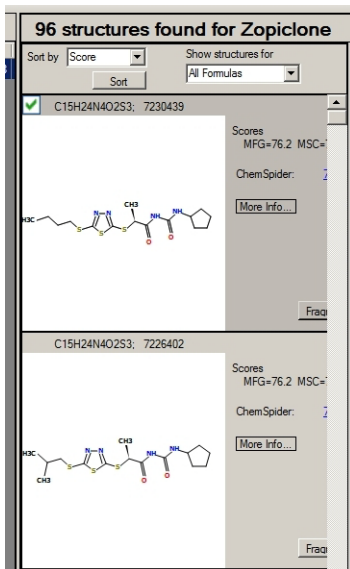


Figure 5 Display of matching candidate structures

To confirm a proposed structure

- 1 Click **Settings** > **Structure Sources**.
- 2 Select and open a molfile.

Step 5. Define structure candidates to correlate to the selected formula and MS/MS spectrum

You can load one file after another. The result is appended in the display. MSC immediately executes the correlation, and then displays the selected structure in the Structural Information pane. The results for each fragment ion are shown in a table above the structures. The results include a value for the percentage of ions explained with a plausible substructure by number and by weighted total ion intensity.

Structure #	Mass	Intensity	Weight(%)	No. of candid.	Best score
1	59.0604	1622.00	0.0	2	85.5
2	98.0844	866.00	0.0	2	73.1
3	99.0916	9783.00	0.5	2	78.8
4	143.0814	6272.00	0.7	5	91.1
5	217.0271	2131.00	0.6	2	34.0
6	217.1076	1540.00	0.4	0	0.0
7	245.0221	106818.00	35.9	1	85.9
8	245.0374	8618.00	2.9	3	56.5
9	245.0523	4826.00	1.6	2	83.2
10	245.0898	1331.00	0.4	0	0.0

Figure 6 MSC results for confirmation of proposed structure

- To sort structures by compatibility score after manually adding structures, click **Sort**.
- To restrict correlation results, click **Settings > Structure Filters** and change the settings in the **Structure Filters** dialog box.

Getting Started

Step 5. Define structure candidates to correlate to the selected formula and MS/MS spectrum

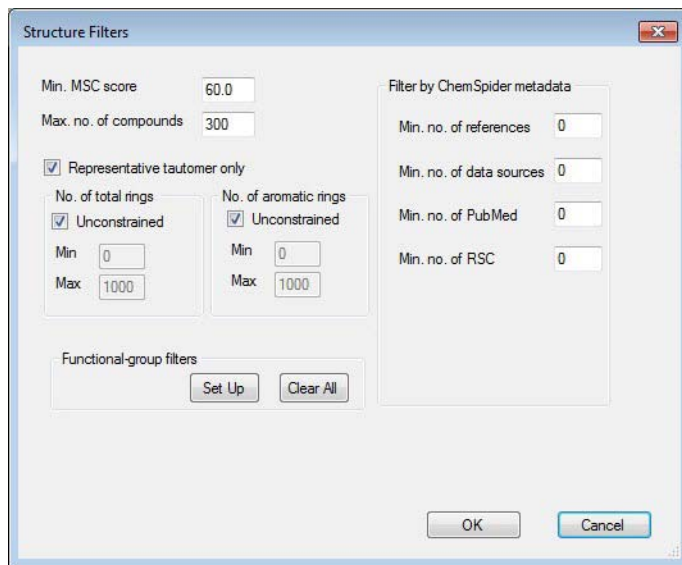


Figure 7 Structure Filters dialog box

To identify a true unknown structure

- 1 Click **Settings > Structure Sources** to open the **Structure Sources** dialog box. Then select a source for structures to correlate.

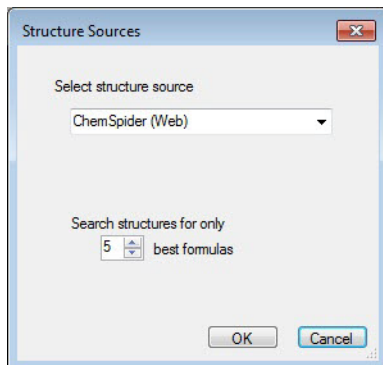


Figure 8 Sources to compare structures for correlation

Step 5. Define structure candidates to correlate to the selected formula and MS/MS spectrum

You can select from:

- ChemSpider compound database on the web (internet connection required)
- Any of the MassHunter Personal Compound Databases (PCD) and/or Libraries (PCDL) in the **D:\MassHunter\PCDL** folder. To use a PCD or PCDL in a different location, select **Other Agilent PCD/PCDLs**.
- SDF files (containing multiple structures)
- Molfile files (containing a single structure)

2 Click **OK** to close the **Structure Sources** dialog box.

3 Click .

MSC retrieves all structures from the selected source that match the selected formula, executes the correlation, and displays all candidate structures in the center pane, sorted by correlation score. The correlation results for the highest scoring structure is shown in the right pane. To see the correlation results for another structure, select it in the Structural Information pane, left side.

4 To restrict the correlation results, click **Settings > Structure Filters**, and set parameters in the **Structure Filters** dialog box. See [Figure 7](#).

The fragment-structure correlation is done automatically on all matching structures. The returned structures are graphically displayed with the highest scoring structure at the top of the list.

MSC displays the total number of structures that are in the database, and the number of structures returned which satisfy the search criteria.

Step 6. Review the candidate structures

- 1 Review the candidate structures that are displayed.

MSC displays the structures that match the correlation criteria, obtained in the previous step. MSC sorts the possible structures by compatibility score and lists the one with the highest score on top.

- 2 Click one of the source hyperlinks displayed to get additional information from the web about the particular compound and structure (see [Figure 5](#) on page 12).

If only a CAS ID is available (e.g. for an entry in a PCD/PCDL database), you are directed to the respective entry on the PubChem website is provided. If a ChemSpider ID is associated with the compound, you are directed to the respective entry on the ChemSpider website.

Step 7. Select a structure to view its substructure assignment

- Click **MSC** next to a structure of interest to review the substructure to fragment ion assignments.

A valid substructure must have the same mass as in the data and be a part of parent compound structure.

MSC displays a table listing each fragment (m/z , intensity, etc) in the data along with the number of possible substructures which has the fragment m/z and is a part of the structure.

The various substructures that match a fragment are highlighted in blue in the structure.

Step 8. Review the possible substructures for each fragment peak

- In the Fragments table, double-click the fragment number (left most column) for a specific fragment to see possible assignments for that fragment.

The location of the fragment on the compound structure for each possible assignment is displayed below the table. The penalty and delta mass and fragment formula are displayed with the structure. An up/down list box allows viewing additional locations of a fragment where it is possible (for example, where a fragment is a mirror image).

Some substructures with high penalty (low scores) are not shown.

Step 9. (Optional) Select another formula

To select another formula, repeat these steps:


- “Step 4. Select a formula for fragment structure correlation analysis” on page 12
- “Step 5. Define structure candidates to correlate to the selected formula and MS/MS spectrum” on page 12
- “Step 6. Review the candidate structures” on page 16
- “Step 7. Select a structure to view its substructure assignment” on page 16
- “Step 8. Review the possible substructures for each fragment peak” on page 17

Familiarization Exercises

The MSC installation disc contains example data files that you can use to get familiar with MSC:

- Proadifen.cef
- Ritonavir C37H48N6O5S2.ms2
- TOX_autoMSMS_01.cef



Exercise 1. Confirm a proposed structure



- 1 Create a new project that contains **Proadifen.cef**:
 - a Click **Project > Create** to open the **Create Project** dialog box.
 - b Type a **Project name**, such as **Exercise1**.
 - c Click **Add files**, and then select **Proadifen.cef**.
 - d Click **OK** to close the **Create Project** dialog box and create the project.
 - e Click .
- 2 Make sure the formula C23H31NO2 is selected as the formula of interest in the upper left pane.
- 3 To add a structure, right-click the first cell, the cell that contains the first row number, of the **Proadifen** row in the **Spectrum List** table, and then select **Add molfile**.
- 4 Select the **Proadifen.mol**. MSC will load the structure, execute the correlation, display the Proadifen structure in the center pane, and display the fragment substructure correlation results in the right pane.
- 5 Click on the different fragment ion rows in the upper half of the right pane and review the assigned substructures. Notice the different penalties which reflect the energy necessary to break bonds resulting in the different substructures.

In the left lower pane, MSC cannot explain several ions in the MS/MS spectrum with a molecular formula that represents a sub formula of the formula proposed for the precursor ion (they are colored grey). Most likely those ions are not fragments of the compound represented by the selected formula, or are chemical noise.

In the upper right pane, ions that are not explained in the proposed structure using the bond-breaking approach are highlighted in a light gray text. The fact that two independent approaches (MFG and MSC) show similar results provides strong confidence that those ions do not belong to the proposed structure.

Exercise 2. Identify true unknowns

- 1 Create a new project and open the file **Proadifen.ccf**:
 - a Click **Project > Create** to open the **Create Project** dialog box.
 - b Type in a name for the project, such as **Exercise2**.
 - c Click **Add files**, and then select **Proadifen.ccf**.
 - d Click **OK** to close the **Create Project** dialog box and create the project.
 - e Click .
- 2 Make sure the formula **C23H31NO2** is selected as the formula of interest in the upper left pane.
- 3 Click **Settings > Structure Sources** to open the **Structure Sources** dialog box. Select **ChemSpider (Web)** as the compound database (Internet connection required).
- 4 Click .

Sometimes when you click  for the first time, “0 structure found C23H31NO2” is returned, which is not correct. Click **Go** again, which usually returns around 60 structures in the center pane.
- 5 Identify the correct structure for Proadifen, which has the ChemSpider ID 4741.
- 6 Click the hyperlink **4741** to access all compound information available in ChemSpider for Proadifen.
- 7 *Optional.* If you own the MassHunter Forensics/Tox PCD or PCDL, select it in from the **Structure Search** drop-down list. If the PCD or PCDL is not in the default folder, click **Other Agilent PCD/PCDL** and browse to its location.
- 8 Click  to get the results.

Familiarization Exercises

Exercise 2. Identify true unknowns

The MassHunter Forensics/Tox PCD or PCDL returns 3 hits, including the Proadifen. The correlation scores between those three are more discriminated than the scores for all the structures in ChemSpider. If you have built a compound database that is more relevant to your specific application (such as one that contains compounds that are of the same class as your unknown), then search the smaller compound database first, before you search ChemSpider.

Exercise 3. Correlate an MS/MS spectrum input via a .ms2 text file

- 1 Open the file **Ritonavir C37H48N6O5S2.ms2**.
The m/z values in the text file were taken from the most abundant fragment ions in an MS/MS spectrum from the literature. The correct formula does not show up as the top hit after molecular formula generation. Look for the correct formula in the upper left table and select it.
- 2 To add a structure, right-click the first cell of the Ritonavir row in the **Spectrum List** table, then select **Add molfile**.
- 3 Select **Ritonavir C37H48N6O5S2.mol**. MSC loads the structure, executes the correlation, displays the Ritonavir structure in the center pane, and displays the fragment substructure correlation results in the right pane.
- 4 *Optional.* To enter a different formula for **Ritonavir**, right-click the first cell of the row for that compound in the **Spectrum List** table, select **Impose formula** to open the **Input Formula** dialog box. Type in the Ritonavir molecular formula (C37H48N6O5S2), and then click **OK**.
- 5 Notice that the formula is displayed in red, and that the fragments are displayed in gray with no subformulas calculated (the MFG is not executed).
- 6 Repeat [step 2](#) and [step 3](#) to review the results.

Exercise 4. Select an MS/MS spectrum for correlation from a CEF file which contains multiple MS/MS spectra

- 1 Open the file **TOX_autoMSMS_01.cef**.
- 2 In the **Spectra Selector** dialog box, select compound 5 with a precursor m/z of 354.2437.
This is the spectrum for Proadifen.
- 3 Follow the steps in “[Exercise 1. Confirm a proposed structure](#)” and “[Exercise 2. Identify true unknowns](#)” to conduct the correlation and review the results.
- 4 *Optional.* Open **TOX_autoMSMS_01.cef** again, and run the correlation on the other 4 MS/MS spectra.

Reference

Basis for MSC

The MSC program is based on the “Automated assignment of high-resolution collisionally activated dissociation mass spectra using a systematic bond disconnection approach” publication from Alastair W. Hill and Russell J. Mortishire-Smith, *Rapid Communications in Mass Spectrometry* 2005; **19**: 3111–3118, published online in Wiley InterScience
<http://onlinelibrary.wiley.com/doi/10.1002/rem.2177/pdf>

Structure of .ms2 file

The “ms2” text file (*.ms2) is an ASCII based text file. You can use Microsoft Notepad to create and edit it. A valid ms2 file must and only have the following lines in the order described here.

- 1 One header line with string **#precursorPeaks *n*** where *n* indicates the number of data lines that follow.
- 2 One or more lines of *m/z* and intensity values. Each line has one pair of *m/z* and intensity values, separated by a space, a tab, or a comma. The number of lines must match the number that follows **#precursorPeaks**.
- 3 One header line with string **#precursorChargeState *n*** where *n* is the charge state. Do not use negative numbers even if the instrument polarity is negative.
- 4 One header line with string **#precursorIonSpecies **M+H**** where **+H** is the ion species. MSC support four different ion species: **+H** (positively charged with protons), **-H** (negatively charged with protons), **+** (positively charged with electrons), and **-** (negatively charged with electrons).
- 5 One header line with string **#productPeakList *n*** where *n* indicates the number of data lines that follow.
- 6 One or more lines of *m/z* and intensity values. Each line has one pair of *m/z* and intensity values, separated by a space, a tab, or a comma. The number of lines must match the number that follows **#productPeakList**.

7 EOF (end of file).

An example of the content of an ms2 file:

```
#precursorPeaks 3
283.1565      1.0
284.1599      0.2
285.1628      0.023
#precursorChargeState 1
#precursorIonSpecies M+.
#productPeakList 15
65.039,4436
66.0457,759
67.0546,2569
69.0701,1289
73.0471,7449
74.0153,1557
74.048,528
75.0246,3138
76.0303,1901
77.0388,7112
78.046,2000
79.0539,2596
80.0577,490
81.0701,1785
83.0495,921
```

www.agilent.com

In This Book

This book contains instructions to install and use the Agilent MassHunter Molecular Structure Correlator (MSC) Software.

Windows 7[®] is registered trademark of Microsoft Corporation in the US and/or other countries.

This guide applies to the Agilent MassHunter Molecular Structure Correlator (MSC) Software B.07.00 or higher, until superseded.

© Agilent Technologies, Inc. 2014

Revision A, September 2014



G3335-90176



Agilent Technologies