Agilent MassHunter Molecular Structure Correlator (MSC) Software

Quick Start Guide

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 20).

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

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

The MSC then uses the selected formula, retrieves one or multiple possible structures from a .mol file, an .sdf file, a MassHunter compound database (PCD, PCDL) 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.

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 for this purpose 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 most likely formula from a local compound database (MassHunter PCD or PCDL) or a web compound database (ChemSpider). The candidates with higher scores are more likely to be the correct structure.

MSC can help to identify the most probable molecular formula 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 a particular formula. 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 to.

Installation

1 Run **MSC.msi** to install the program.

2 The files are installed in **C:\Program Files (x86)\Agilent\MassHunter\Workstation\MSC**.

3 Copy example data (.cef files) from the **Data** folder on the installation CD to a location of choice on your hard drive.

4 If you have not installed PCDL Manager B.04.00 (or higher), or you are not running Windows 7, install the SqlServer Compact Edition 3.51 (SSCE), located in the **SSCERuntime** folder (-x86 version).

Getting Started

The MSC application has three main panes.

- The left pane displays the MFG result (see "Step 3. Review molecular formulas for the compound" on page 9).
- The center pane shows the candidate structures of the parent compound (see "Step 6. Review the candidate structures" on page 15
- The right pane is used to review the substructure assignments (see "Step 8. Review the possible substructures for each fragment peak" on page 15).

Figure 1   User Interface of Molecular Structure Correlator

At the top are three toolbar icons:

- Click ⬜ to open data files.
- Click ⬜ to specify how the fragment spectrum is acquired from the instrument.
- Click ⬜ to specify chemical information of the compound.

Step 1. Load fragment spectrum

1 Click 📁 in the main toolbar.

2 Specify which kind of file to open (*.cef or *.ms2).
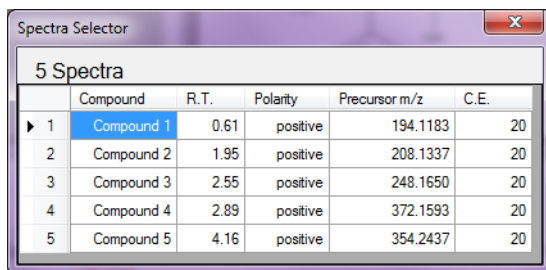
The .cef files are created by the MassHunter Qualitative Analysis program. The format for the .ms2 file is specified in "Structure of .ms2 file" on page 20.

3 If the .cef file contains more than one fragment spectra, then in the **Spectra Selector** dialog box, double-click the spectra number (left most column) of one spectrum to select it.

Spectra Selector					
5 Spectra					
	Compound	R.T.	Polarity	Precursor m/z	C.E.
▶ 1	Compound 1	0.61	positive	194.1183	20
2	Compound 2	1.95	positive	208.1337	20
3	Compound 3	2.55	positive	248.1650	20
4	Compound 4	2.89	positive	372.1593	20
5	Compound 5	4.16	positive	354.2437	20

Figure 2 Spectra Selector for .cef files

Step 2. Enter compound chemistry and data quality information

1 Click **Settings > Spectra Processing Parameters** to open the **Spectra Processing** dialog box.
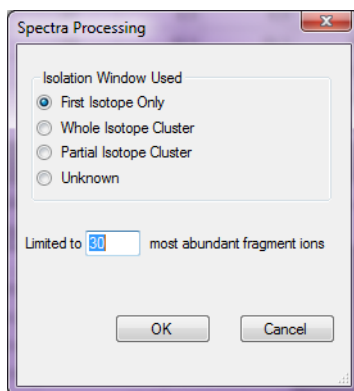
Figure 3   Spectra Processing Parameters

1 Complete the dialog box and click **OK**. Refer to online Help for more information.

The options for **Isolation Window Used** 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).

2 If you need to change the MFG parameters, click ▦ in the toolbar.

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

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

Figure 4   Setup MFG Parameters

Note that whenever you click **OK** in either the **Spectra Processing** or the **Composition Info** dialog boxes, MSC clears all the existing results (if any) and automatically recalculates the MFG results for the current data.

Step 3. Review molecular formulas for the compound

The upper pane lists the candidate formulas that MSC calculated for the compound.

- If the formula you want is not in the table, type it in the text box under **Compound formula**, and then click **Add**. The new formula, if valid, is displayed in the table.

  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, is not run.

  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.

Figure 5   Molecular Formula Generation

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 be correlated to the accurate mass MS/MS spectrum, depending on your application.
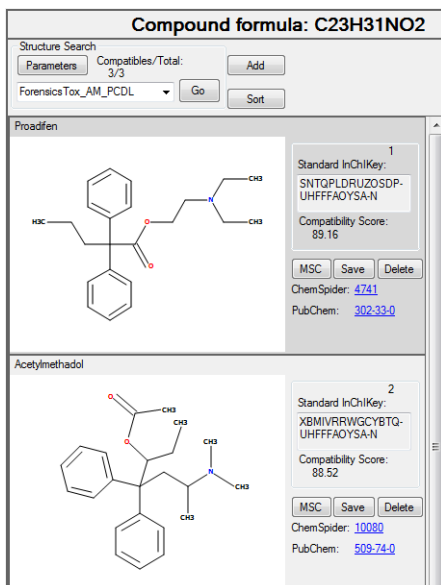
Figure 6   Display of matching candidate structure

To confirm a proposed structure

1 Click **Add** to the right of **Structure Search**.

2 Select and open a molfile when prompted.

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 center pane and the results in the right pane. The upper pane lists the results for each fragment ion including the percentage of ions explained with a plausible substructure by number and by weighted total ion intensity.

Figure 7 MSC results for confirmation of proposed structure

3 To sort structures by compatibility score after manually adding structures, click **Sort**.

4 To restrict correlation results, click **Parameter** under **Structure Search** and change the settings in the **Search Parameters** dialog box.
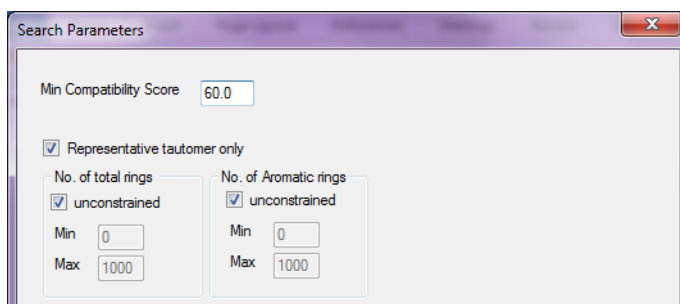
Figure 8 MSC Search Parameters

To identify a true unknown structure

1 From the drop-down list under **Structure Search**, select a source for possible structures to correlate.
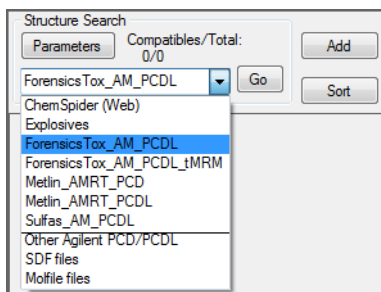
Figure 9   Sources to compare structures for correlation

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) or .mol files (containing a single structure).

2 Click **Go**.

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 center pane.

3 To restrict the correlation results, click **Parameters** and set parameters in the **Structure Search Parameters** dialog box. See Figure 8.

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

To the right of **Parameters** and **Open** buttons, 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 hyperlinks displayed to get additional information from the web about the particular compound and structure (see Figure 6 on page 11).

If only a CAS ID is available (e.g. for an entry in a PCD/PCDL database), a hyperlink to the respective entry on the PubChem website is provided. If a ChemSpider ID is associated with the compound, the "ChemSpider" hyperlink will connect 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 10
- "Step 5. Define structure candidates to correlate to the selected formula and MS/MS spectrum" on page 11
- "Step 6. Review the candidate structures" on page 15
- "Step 7. Select a structure to view its substructure assignment" on page 15
- "Step 8. Review the possible substructures for each fragment peak" on page 15

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 Open the file **Proadifen.cef** file.

2 Make sure the formula C23H31NO2 is selected as the formula of interest in the upper left pane.

3 In the center pane, click **Add**.

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, MSC can also not explain many of those ions into the proposed structure using the bond-breaking approach (they are colored grey). 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 Open the file **Proadifen.cef**.

2 Make sure the formula **C23H31NO2** is selected as the formula of interest in the upper left pane.

3 In the center pane, select **ChemSpider (Web)** as the compound database (internet connection required).

4 Click **Go**.

Sometimes when you click **Go** 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 Under **Compatibles/Totals**, check how many unique structures are returned from ChemSpider for the selected formula, and how many of those resulted in a correlation score above the selected value in **Parameters** (default 60).

6 Identify the correct structure for Proadifen, which has the ChemSpider ID 4741.

7 Click the hyperlink **4741** to access all compound information available in ChemSpider for Proadifen.

8 *Optional.* Select another structure that was returned, and click **MSC** next to the structure to display and review the results.

9 *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.

10 Click **Go** to get the results.

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 In the center pane, click **Add**.

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.* Type in the Ritonavir molecular formula (C37H48N6O5S2) for **Compound formula** in the upper left pane, and then click **Add**.

5 Notice that the formula is displayed in red, and that the fragments are displayed in grey 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/rcm.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).

The next page contains 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.

G3335-90126

Agilent Technologies