MassBank User's Manual (in English)

(Version 1: 2020/01/14)

- This manual is the primary version and will be updated with the improvement of MassBank search tools.
- MassBank Record Format is downloadable from
  < https://github.com/MassBank/MassBank-web/blob/master/Documentation/MassBankRecordFormat.md >.
- We welcome questions and comments on this manual and MSSJ MassBank database service. Please contact us by massbank@mssj.jp.

1. Access to MSSJ MassBank.jp server (http://massbank.jp/)

Figure 1. Home page of MSSJ MassBank.jp server

2. Jump to （1）“Spectral Search” pages or （2）“Statistics of MassBank Data” page.

Figure 2. Home page of MSSJ MassBank.jp server.

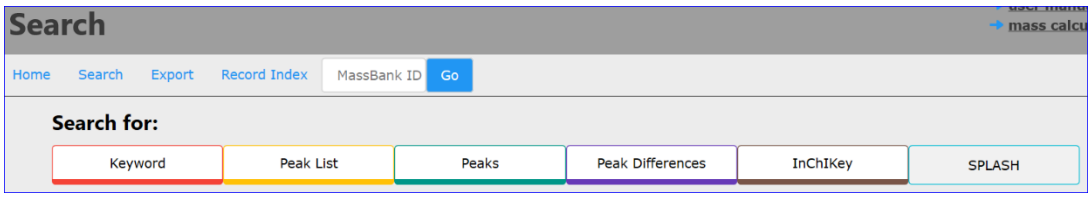
(1) To “Spectral Search” page
Click Search.

(2) To “Statistics of MassBank Data” page
Click Record Index.

3. Search tools for mass spectral data

Currently six tools are available (as of January 2020).

Figure 3. Panel for selecting search tools.



Note 3: Search tools require the users to define (1) search query and (2) search target.

(1) Search query is chemical name, a mass spectrum, a set of spectral peaks, numerical values, or text. See Figure 3.1 as an example.

(2) Search target is a set of mass spectra that are selected from MassBank by specifying the following "Mass Spectrometry Information"; Ionization, Instrument type of mass analyzer, MS type, and Ion mode. See Figures 3.1.1 and 3.1.2 as examples.
About the setting of "Ion Mode" in "Mass Spectrometry Information", see Note 3.1.3 (Page 6).

Brief introduction to search tools and their functions.

3.1. "Keyword" search

This tool finds all the mass spectra that analyzed a chemical substance specified by the chemical name. Search query is the name of the chemical substance or a part of the name. Chemical substance usually has two or more different chemical names. Nevertheless, most contributors to MassBank give only one chemical name to each chemical substance. "Keyword" is often different from the chemical names given by contributors. In this case, a "Keyword" search will not hit the corresponding mass spectra. If you know the chemical structure of the chemical substance, search by InChIKey of the chemical substance gives you comprehensive results (see "3.5. InChIKey" search).

3.2. "Peak List" search

This tool finds the mass spectra most similar to the query mass spectrum. Search query is a whole or part of a mass spectrum; that is, a list of peaks expressed by the *m/z* and relative intensity values. Similarity of two spectra is evaluated by considering the number of the peaks matched in the m/z value with similar intensity value. Similarity between the query mass spectrum and every one of the mass spectra in the search target is pairwise evaluated as similarity score. This search tool outputs the mass spectra in the order of the score. "Peak List" search and "Peaks" search are different. See "3.3 Peaks search".

3.3. "Peaks" search

This tool finds the mass spectra that observed all the peaks specified by a set of *m/z* values. Search query is a list of the *m/z* and intensity values, though intensity values are ignored during the search. On the contrary, "3.2. Peak List search" does not require that all the peaks specified by search query are observed in the

target spectra. And “3.2. Peak List search” considers the intensity values.

3.4. “Peak Differences” search (At present this tool does not work.)

This tool finds the mass spectra that observed the neutral losses specified by one or more molecular formulae as search query.

3.5. “InChIKey” search

This tool finds the mass spectra that analyzed the chemical substance identified with the InChIKey. InChIKey, consisted of 27 characters, is a hashed counterpart of InChI that is a text identifier for chemical substances. Search query is InChIKey with 27 characters (the “complete” InChIKey) or with the first 14 characters (the “parts” InChIKey). While the “parts” InChIKey specifies all the stereoisomers of the specified atom connectivity, the “complete” InChIKey identifies one stereoisomer. Mass spectrometry gives the same spectrum to all the stereoisomers of the same atom connectivity. Therefore, the “parts” InChIKey provides comprehensive search results (see “3.1. Keyword search”).

3.6. “SPLASH” search

This tool finds the mass spectrum specified by SPLASH(*) that is an unambiguous, database-independent spectrum identifier.

(*) Wohlgemuth G, Mehta SS, Mejia RF, Neumann S, edrosa D, Pluskal T, Schymanski EL, Willighagen EL, Wilson M, Wishart DS, Arita M, Dorrestein PC, Bandeira N, Wang M, Schulze T, Salek, Steinbeck RMC, Nainala VC, Mistrik R, Nishioka T & Fiehn O, “SPLASH, a hashed identifier for mass spectra”, *Nature BioTech*, **34** (11), 1099-1101 (2016). doi:10.1038/nbt.3689.

3.7. “MassBank Record ID” search

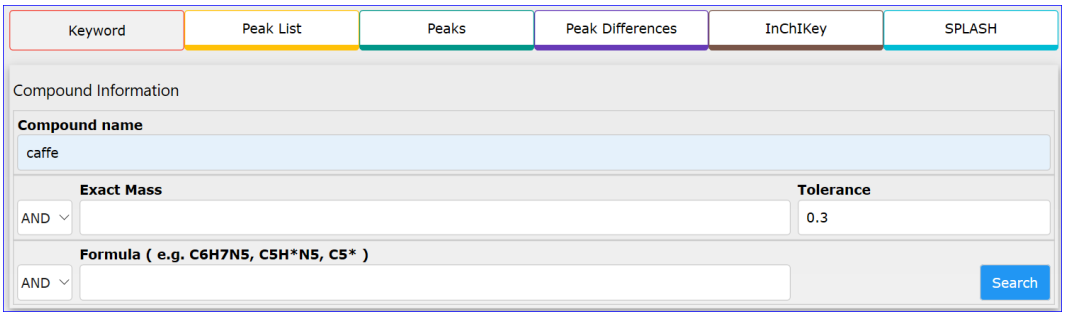
This tool gets the mass spectrum specified by the MassBank Record ID as search query.

3.1. “Keyword” Search

Function: This tool finds all the mass spectra that analyzed a chemical substance specified by the chemical name (mandatory) with additional chemical information (optional).

[1] Crick to select “Keyword” on the panel for selecting search tools (Figure 3).
[2] Panel appears for “Keyword” search query (Figure 3.1).
Search query is the chemical name in the “Compound name” cell (mandatory). Monoisotopic “Exact Mass” (theoretical value) with “Tolerance” and molecular “Formula”, are additional information (optional). "Tolerance" is the deviation of the observed mass from "Exact mass".
[3] Define search target (Figures 3.1.1 and 3.1.2).
See Notes 3.1.1, 3.1.2, and 3.1.3 (pages 5, 6).
[4] Crick “Search” button in the lower right corner of Figure 3.1 to start the search.
[5] “Keyword” search results are summarized in “Quick Search Results” panel (Figures 3.1.4a and 3.1.4b.).
See Notes 3.1.4, 3.1.5, and 3.1.6 (pages 6, 7).

Figure 3.1. Panel for the input of “Keyword” search parameters.

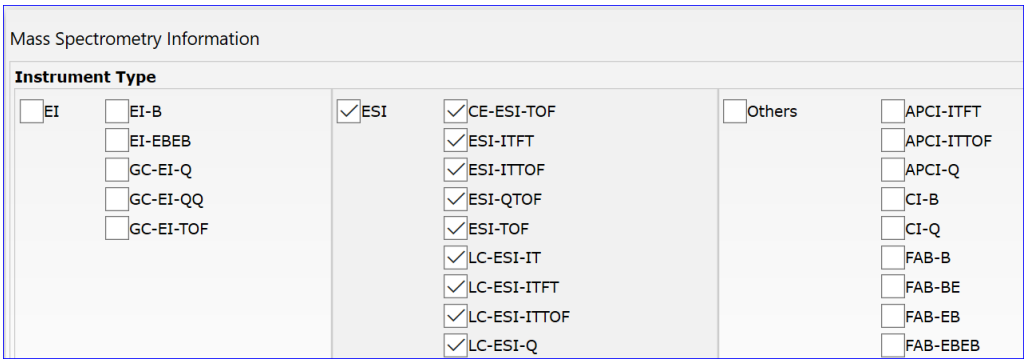


Tips for using “Keyword” search.

When you input “caffe”, as an example, in the “Compound name” cell, this tool finds the mass spectra that analyzed such chemical substances as “Caffeic acid” and “Dihydrocaffeic acid”. In addition, the mass spectra analyzed “3,4-Dimethoxycinnamic acid” are selected because the alias of the substance “Dimethyl caffeic acid” is given to these spectral data. Wild card “*” is accepted as a part of molecular “Formula”.

Note 3.1.1: “Mass Spectrometry Information” panel (Figure 3.1.1) is displayed below the “Input of search parameter” Panel. Specify “Ionization method” and “Type of mass analyzer” in the Instrument Type by selecting boxes by check. These are mandatory information.

Figure 3.1.1. Use the check boxes to specify the ionization method and the type of mass analyzer.



Note 3.1.2: Panel “MS Type” and “Ion Mode” (Mandatory)

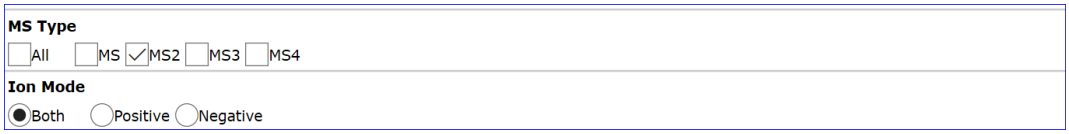
The panel (Figure 3.1.2) might be displayed out of your PC monitor. Don’t leave the Mass Spectrometry Information panel without scrolling up the panel and checking the boxes of “MS Type” and “Ion Mode” (Figure 3.1.2).

Figure 3.1.2. Use the check boxes and circles to specify the “MS Type” and “Ion Mode”.

Note 3.1.3: Important notice about the “Ion Mode” setting.

The same search might be repeated to the same search target. In the “Mass Spectrometry Information” panel, the check in the boxes of “Ion Mode” is automatically reset to “Ion Mode = Both” after every search. Confirm the check of “Ion Mode” before every search!

Figure 3.1.3. Check box of “Ion Mode” is automatically reset to Ion Mode = Both.



Note 3.1.4: “Keyword” search results are summarized in “Quick Search Panel” (Figure 3.1.4a and b). In this example, search query is Keyword = “caffe”, and search target is Instrument Type = “ESI”, MS Type = “MS2”, Ion Mode = “Positive”.

Figure 3.1.4a. Upper panel summarizes the value of search parameters (= “Compound Name”) and “Mass Spectrometry Information” (= “Instrument Type”, “MS Type”, “Ion Mode”).

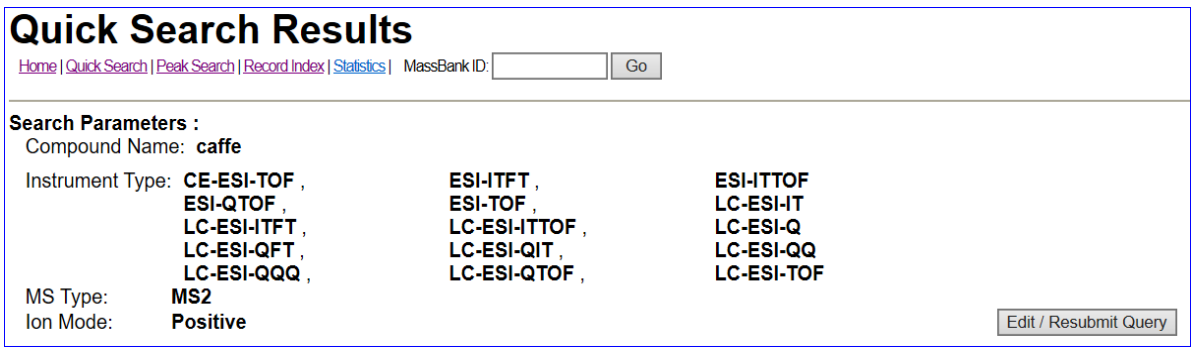
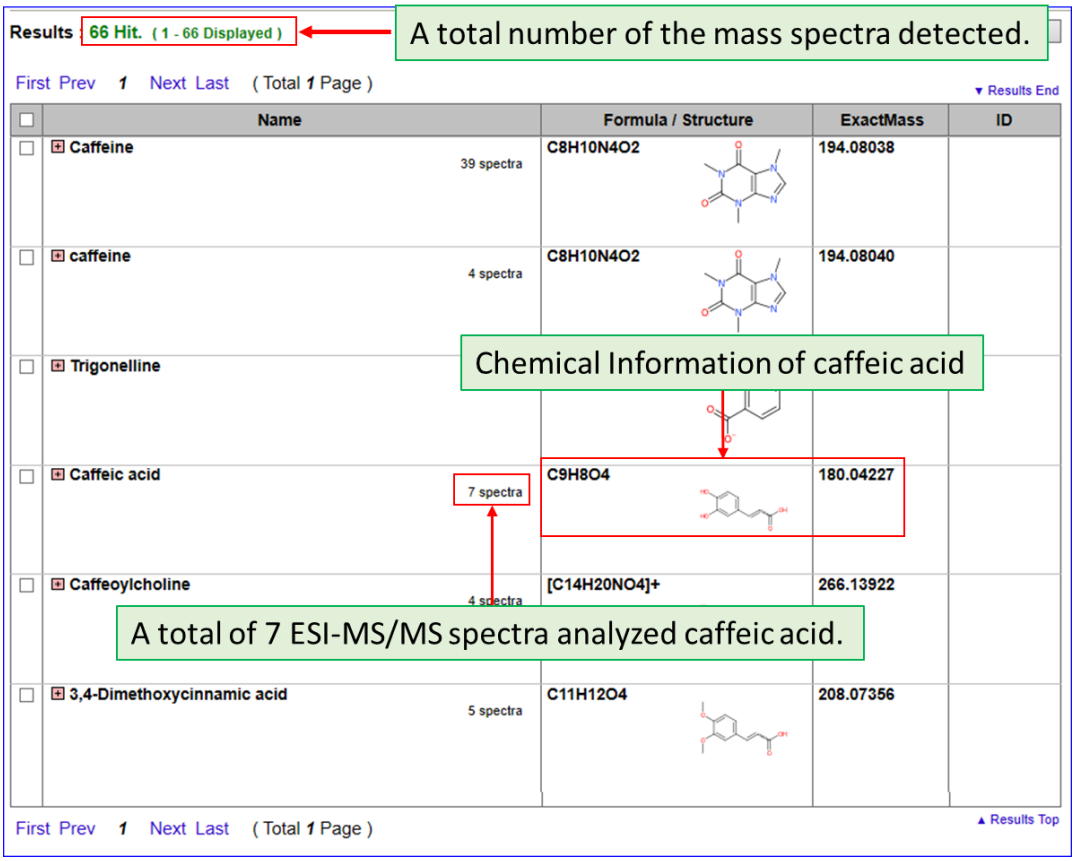
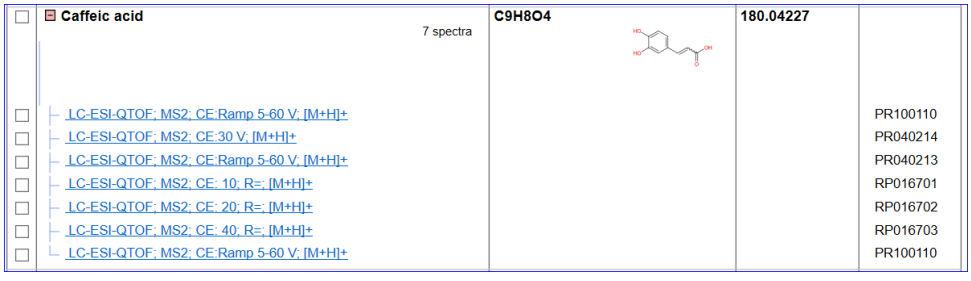


Figure 3.1.4b. Lower panel reports that a total of 66 ESI-MS/MS spectra analyzed 6 “caffe” related chemical compounds were found by the Keyword = “caffe” search. Among them, the number of the ESI-MS/MS spectra that analyzed caffeic acid is seven.



Note 3.1.5: Crick the red box ■ ahead of “Caffeic acid” to expand the list of seven mass spectra (Figure 3.1.5). To expand all the lists detected by the tool, crick Open All Tree located on the upper right corner of the lower panel.

Figure 3.1.5. The list of seven mass spectra analyzed “Caffeic acid”. Record titles of seven ESI-MS/MS spectra are shown on the second column. Eight characters in the rightmost column show MassBank Record ID of mass spectra. Crick “LC-ESI-QTOF; MS2; CE: Ramp 5-60 V; [M+H]+” to display MassBank record PR100110.



Note 3.1.6: “Keyword” is often different from the chemical names given by contributors. In this case, a “Keyword” search will not hit them. If you know the chemical structure of the “Keyword” substance, search by InChIKey instead of “Keyword” gives you comprehensive search results (see “3.5. InChIKey Search”, page 13).

3.2 “Peak List” Search

Function: This tool finds mass spectra most similar to search query which is a whole or part of a mass spectrum; that is, a list of peaks expressed by the *m/z* and relative intensity values.

[1] Crick to select “Peak List” on the panel for selecting search tools (Figure 3).
[2] Panel appears for “Peak List” search query (Figure 3.2.1).
Input a list of peak data in the “Peak data” cell (mandatory). A list of peak data either in NIST format or in one peak data/line format is allowed. Examples of the two data formats are given in “Example 1” and “Example 2” buttons in the lower right corner of the panel. See Note 3.2.1 (page 9) about “Cutoff threshold of relative intensities”. “Number of results” defines the number mass spectra shown in the hit list on the lower panel of “Quick Search Results” (Figures 3.2.3 and 3.2.4).
[3] Define search target (mandatory). See Notes 3 (page 3), 3.1.1, 3.1.2, and 3.1.3 (pages 5, 6).
It is recommended that “Mass Spectrometry Information” of the search target is the same or similar to that of the query mass spectrum.
[4] Crick “Search” button in the lower right corner of Figure 3.2.1 to start the search.
[5] “Quick Search Results” panel reports “Peak List” search results (Figures 3.2.2, 3.2.3). See Notes 3.1.5 (page 7) and 3.2.1.

Figure 3.2.1. Panel for the input of “Peak List” search parameters. A list of peak data in the NIST format are copied from Example 1 and pasted in the “Peak Data” cell.

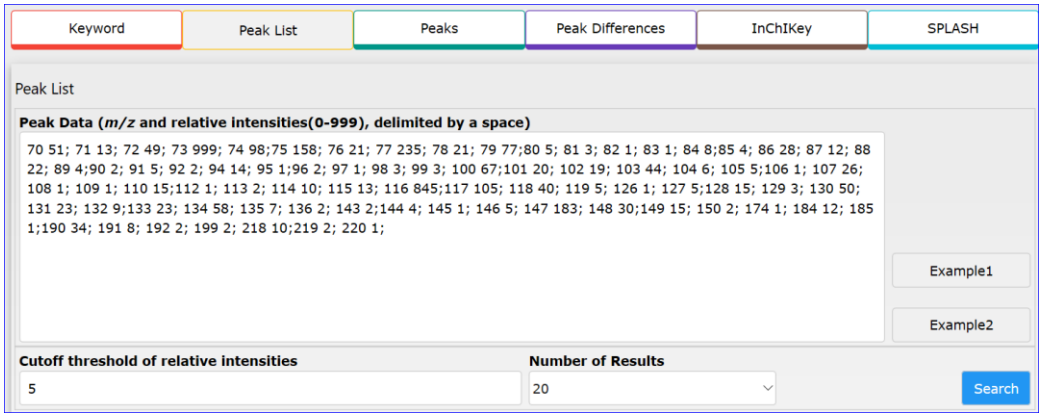


Figure 3.2.2. Upper panel of “Quick Search Results” shows the search query mass spectrum.

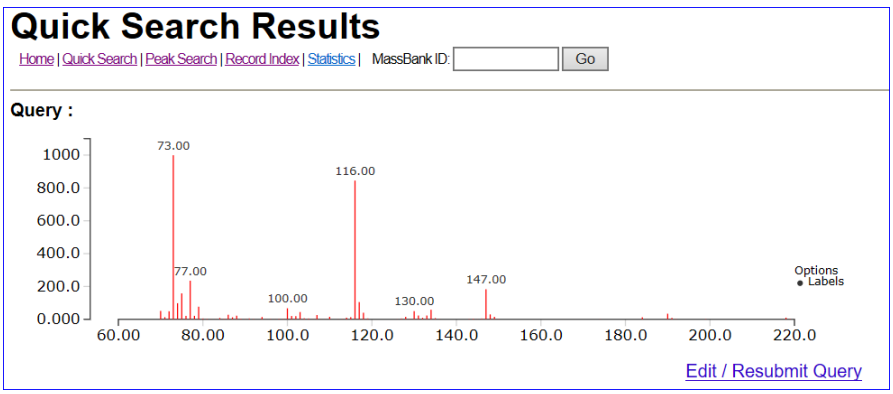
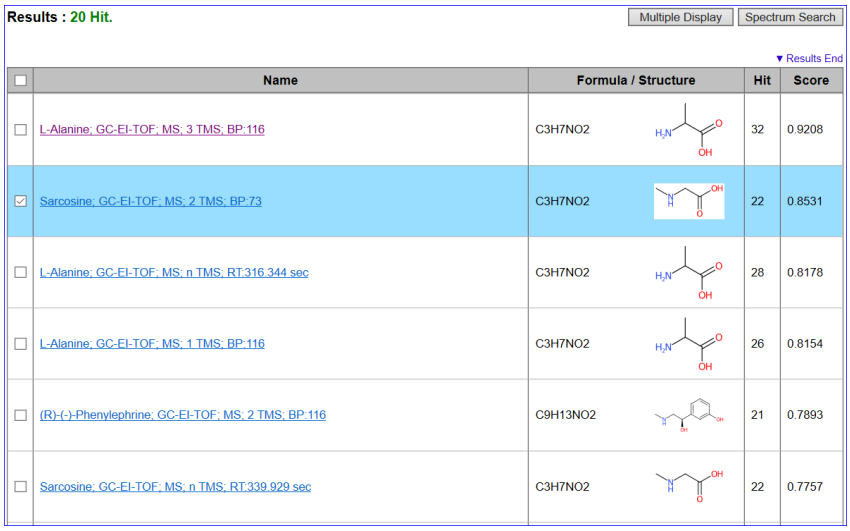
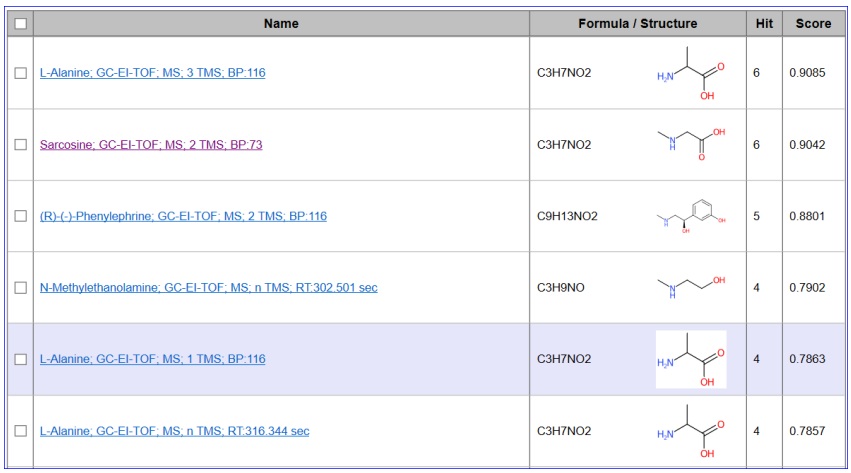


Figure 3.2.3. Lower panel of “Quick Search Results” summarizes the search results by “Peak List” at the cutoff threshold 5. Column “Name” is the title of the MassBank records. Column “Hit” is the number of the peaks that matched to the query peaks on the *m/z* basis. Column “Score” is the similarity score ($0 \leq$ score ≤ 1).



Note 3.2.1. If your query peak list is long and the search target contains many mass spectra, you must wait patiently before the search result is displayed. It might take nearly an hour to get the search results. If you try to get search results while search tool is running, search will be stopped without any results. To reduce the time required for the search, (1) increase the value of the "Cutoff threshold for relative intensity" (Figure 3.2.1), (2) reduce the number of "Device Type" and "MS Type" checks (Figure 3.1.1), and (3) make sure that the "Ion Mode" is not "Both" (Figure 3.1.2).
”Cutoff threshold of relative intensities” removes peaks whose peak intensity is less than the threshold from search query and search target. Default threshold is 0.5 % in MassBank. The 5 % cutoff threshold significantly reduces the time required for the search. When Figure 3.2.4 (Cutoff threshold = 50) is compared with Figure 3.2.3 (Cutoff threshold = 5), the number of the matched peaks in “Hit” and the similarity “Score” are reduced. However, the similarities in the detected compounds are not significantly affected.

Figure 3.2.4. Lower panel of “Quick Search Results” summarizes “Peak List” search results at the cutoff threshold = 50. Other parameter values are the same to those in Figure 3.2.3.



Tips: Zoom-in on mass spectrum.

To zoom in on the mass spectrum (Figure 3.2.5) in the range of *m/z* 50-250, place cursor on the *m/z* axis and draw a rectangle with the base length from *m/z* 50 to *m/z* 250 and almost the same hight as the top peak (Figure 3.2.6). To restore the mass spectrum to its original size, click the cursor anywhere on the enlarged mass spectrum. To show the *m/z* and relative intensity values of a peak, place the cursor on the peak.

Figure 3.2.5. Mass Spectrum (MassBand Record ID: PR010146).

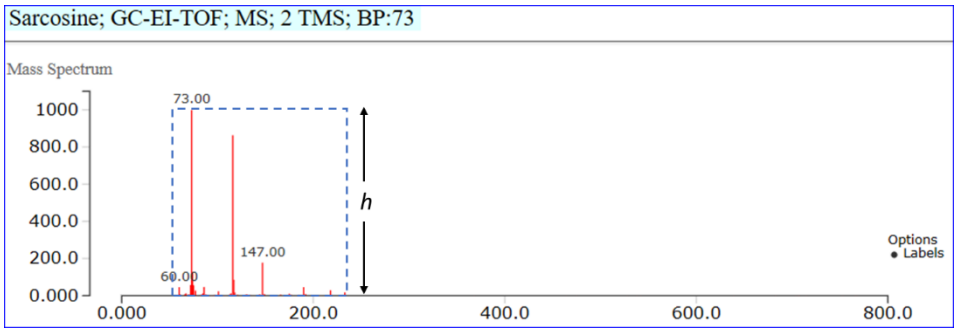
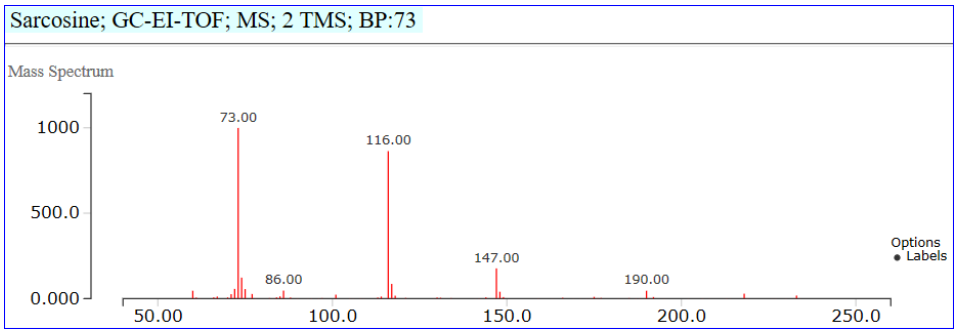


Figure 3.2.6. Mass spectrum enlarged in the range of *m/z* 50 to *m/z* 250.



3.3. “Peaks” Search

Function: This tool finds the mass spectra that observed all the peaks that search query specified by the *m/z* value. “Peaks” Search and “Peak List” Search sound similar, but quite different in the search function.

[1] Crick to select “Peaks” on the panel for selecting search tools (Figure 3).

[2] Panel appears for “Peaks” search query (Figure 3.3.1).

Input *m/z* values in the “m/z” cells (mandatory). The *m/z* values specify the peaks as the search query. When you input the molecular formulae (without “+” and “-“) in the “Formula” cells (Figure 3.3.3), the calculated *m/z* values are automatically input into the “m/z” cells.

The value of “Rel. Intensity” cell at the bottom panel specifies the minimum relative intensity of the spectral peaks in the search target. The default value of 100 indicates that spectral peaks with relative intensity less than 100/999 are excluded from the search target.

"Tolerance" is the deviation of the observed *m/z* of a target peak from the "m/z" of a search query peak. When the deviation is smaller than “Tolerance”, “Peaks” search judges that the two peaks match.

[3] Define search target (mandatory). See Notes 3 (page 3), 3.1.1, 3.1.2, and 3.1.3 (pages 5, 6).

It is recommended that “Mass Spectrometry Information” of the search target is the same or similar to that of the query mass spectrum.

[4] Crick “Search” button in the lower right corner of Figure 3.3.1 or 3.3.3 to start the search.

[5] “Peaks” search results are summarized in “Quick Search Results” panel (Figures 3.3.2a and 3.3.2b). See Notes 3.1.5 (page 7) and 3.2.1 (page 9).

Figure 3.3.1. Upper panel for the input of "Peaks" search parameters. This is an example: Positive ion, *m/z* 94.04, is a search query. Rel.Intensity = 100. Search Criterion is Tolerance = 0.1.

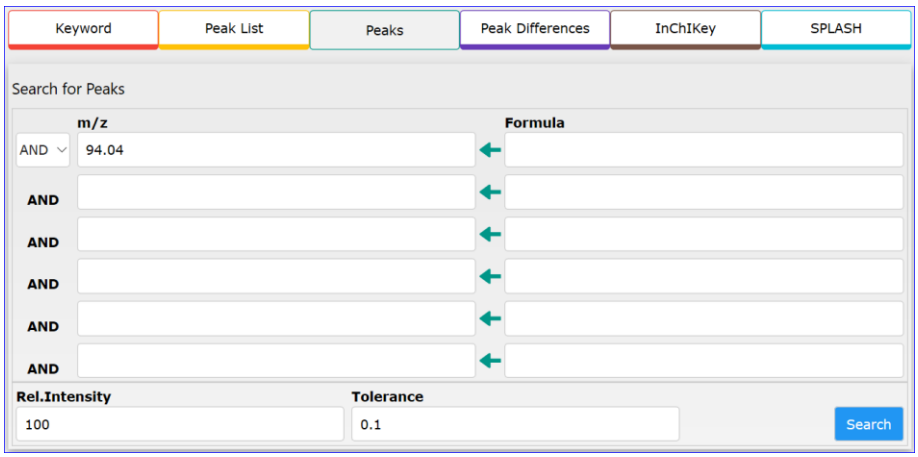


Figure 3.3.2a. Upper panel of "Peaks Search Results" shows the search target.

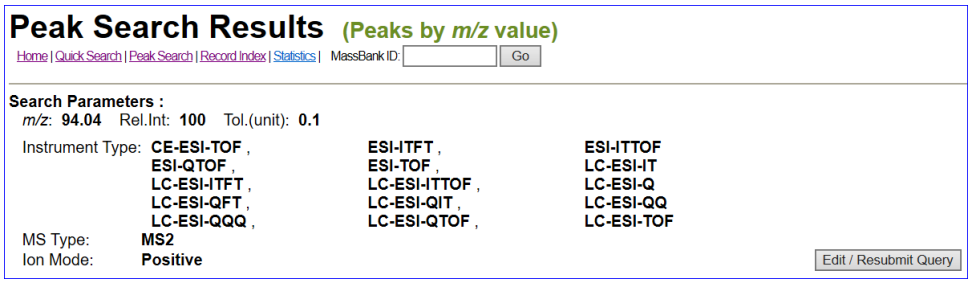
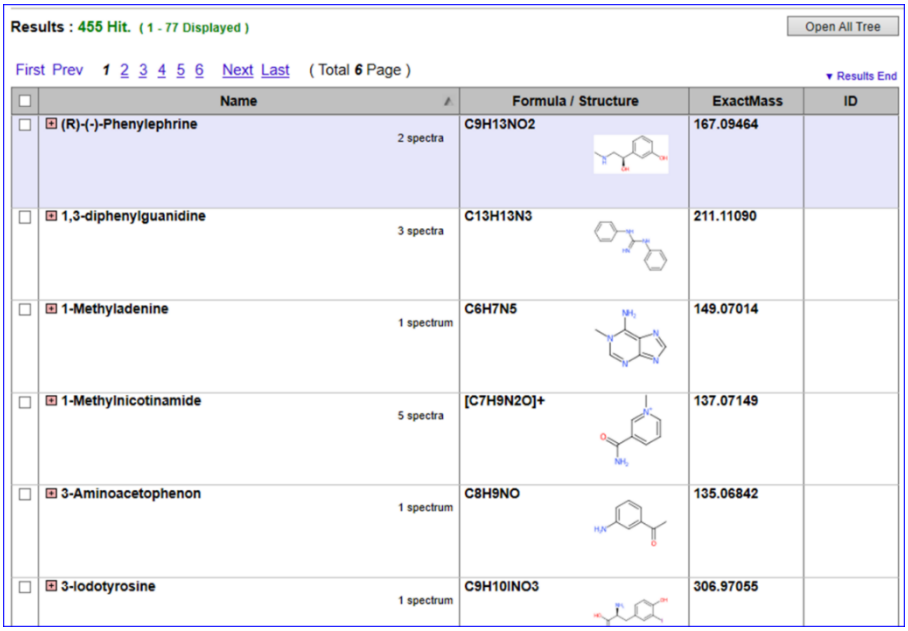


Figure 3.3.2b. Lower panel of "Peaks Search Results" reports that a total of 455 mass spectra that observed the peak *m/z* = 94.04 by analyzing 129 chemical substances.

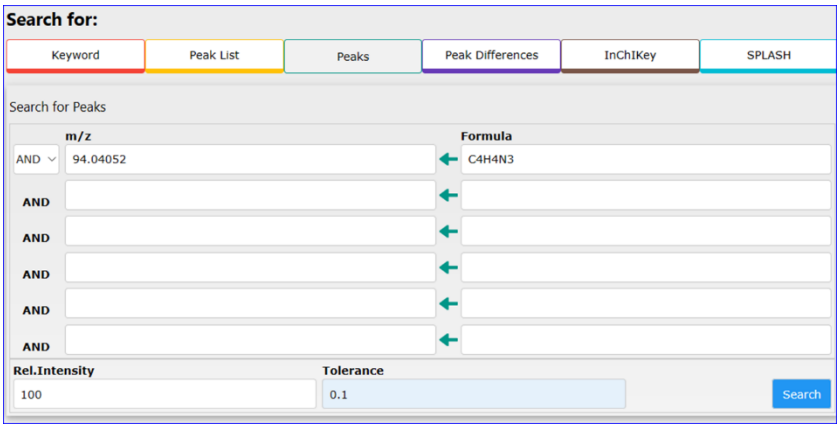


Tips: What can this search tool do for you?

Suppose you want to identify an unknown chemical substance by mass spectrometry. The highest intensity peak was found at *m/z* 94.04 in the mass spectrum. Knowing which chemical substances give the peak at *m/z* 94.04 in the mass spectra can be very helpful in identifying the unknown chemical substance. In such a case, “Peaks” search must be useful.

Current “Peaks” search has a problem to be resolved. Such a problem is shown by an example. When an adenine derivative was analyzed by ESI-MS/MS, a peak at *m/z* 94.04 with higher intensity was observed. Based on studies on the fragmentation of other adenine derivatives, the *m/z* 94.04 peak was supposed as $C4H4N3^+$. “Peaks” search was applied to study whether $C4H4N3^+$ was commonly observed in the mass spectra of hetero unsaturated cyclic compounds ($N \geq 3$) (Figure 3.3.3). However, search result summary is the same to Figure 3.3.2a, b. These search results were unexpected.

Figure 3.3.3. Upper panel for the input of “Peaks” search parameter; $C4H4N3^+$.



The search results summary (Figure 3.3.2b) shows that the chemical compounds containing less than 3 N atoms and more than one O atom are found in addition to those containing more than 3 N atoms. In the “Peaks” search, seach query is *m/z* values even when molecular formulae were input in the search parameters. It is difficult to distinguish between $C4H4N3^+$ (*m/z* 94.0400) and $C6H6O^+$ (*m/z* 94.0413) by the numerical value 94.04. Even when more than two *m/z* values were given as query parameters, “Peaks” search returned hetero unsaturated cyclic compounds containing multiple N atoms and those containing multiple O atoms. This should be noted for mass imaging.

Improvement of “Peaks” search is necessary.

The above "peak" search problem can be easily remedied by adding a refinement. Refinement with molecular formula is necessary. “Formula” in Figure 3.2 is a useful refinment. For example, “>N3*”.

3.4. “Peak Differences” Search (Currently service unavailable)

Function: This search finds the mass spectra that observed the neutral losses specified by molecular formulae.

This search function is currently not working properly. We will investigate the cause and fix it.

3.5. “InChIKey” Search

Function: This tool finds the mass spectra that analyzed the chemical substance identified with InChIKey. InChIKey, consisted of 27 characters, is a hashed counterpart of InChI that is the text identifier for a chemical substance based on the chemical structure. This search is more comprehensive than the “keyword” search (see Note 3.1.6, page 7).

[1] Crick to select “InChIKey” on the panel for selecting search tools (Figure 3).

[2] Panel appears for “InChIKey” search query (Figure 3.5.1).
Search query is InChIKey with 27 characters (the “complete” InChIKey) or with the first 14 characters (the “parts” InChIKey). The first 14 characters specifies the atom connectivity regardless of the stereochemistry.
While the “parts” InChIKey specifies all the stereoisomers of the specified atom connectivity, the “complete” InChIKey specifies one stereoisomer. Mass spectrometry gives the same spectrum to all the stereoisomers of the same atom connectivity. Therefore, the “parts” InChIKey provides comprehensive search results.

[3] Define search target (mandatory). See Notes 3 (page 3), 3.1.1, 3.1.2, and 3.1.3 (pages 5, 6).

[4] Crick “Search” button in the lower right corner of Figure 3.5.1 to start the search.

[5] “InChIKey” search results are summarized in “InChIKey Search Results” panel (Figure 3.5.2a and 3.5.2b). See Notes 3.1.4 and 3.1.5 (page 6, 7). Figures 3.1.4a, and 3.1.4b.

Figure 3.5.1. Panel for “InChIKey” search parameters. In this example, search query is the first 14 characters of the InChIKey of 3-Methylheptyl bromide (“PYETWEQSEZLJIH”).
Search target is “Instrument Type = EI and Others”, “MS Type = MS”, “Ion Mode = Positive”

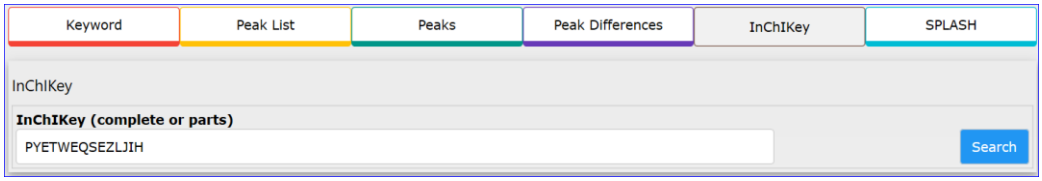


Figure 3.5.2a. Upper panel of “InChIKey Search Results” shows the search target.

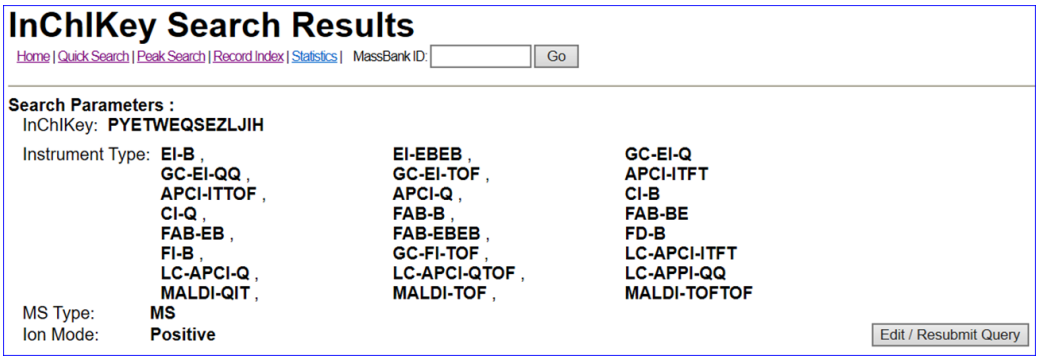
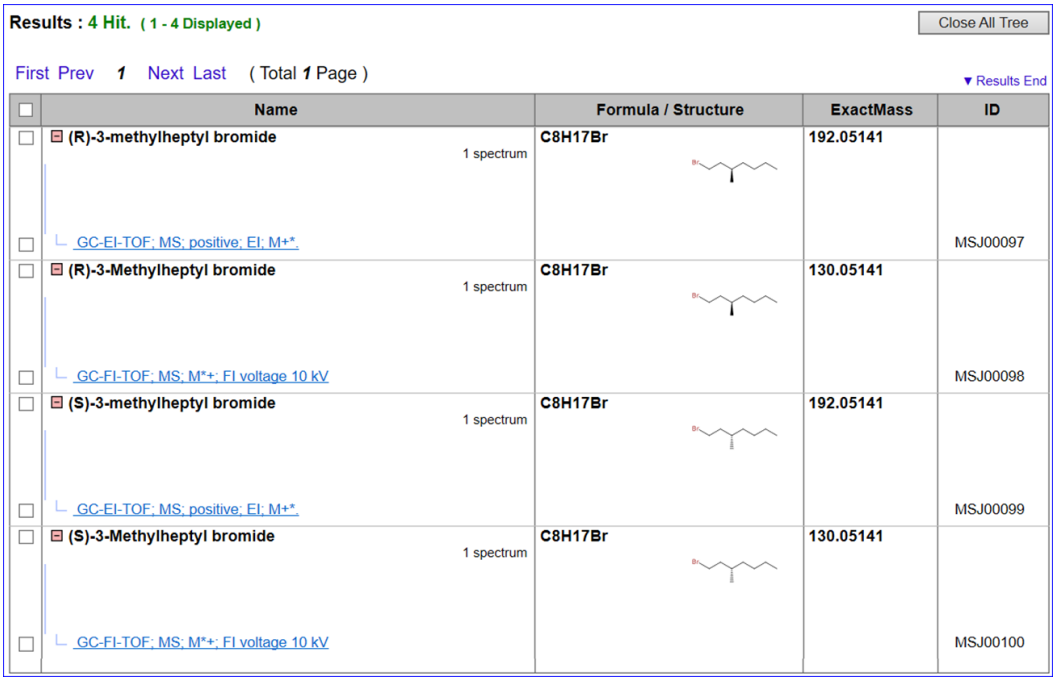


Figure 3.5.2b. Lower panel of “InChIKey Search Results reports that a total of 4 mass spectra analyzed 2 enantiomers of 3-methylheptyl bromide (R and S isomers) by EI-MS and FI-MS.



3.6. “SPLASH” Search

Function: This search is a tool that uses a certain mass spectrum as a query to search for exactly the same mass spectrum (the number of peaks, their *m/z* and relative intensity values are all the same). SPLASH expresses a mass spectrum (*m/z* and strength values) by a hash function.

[1] Crick to select “SPLASH” on the panel for selecting search tools (Figure 3).

[2] Panel appears for “SPLASH” search query (Figure 3.6.1). Input a splash as search query.

[3] Define search target. When you know nothing about the mass spectrometry method of the query mass spectrum, specify all the mass spectra in MassBank as search target (Figure 3.6.2).

[4] Crick “Search” button in the lower right corner of Figure 3.6.1 to start the search.

[5] Summary of “SPLASH” search results is displayed in “Splash Search Results” panel (Figure 3.6.3).

Figure 3.6.1. Upper panel for the input of “SPLASH” search parameters. Example is “SPLASH = splash10-0ukl-0497000000-83f8c283cf428005ba2c”

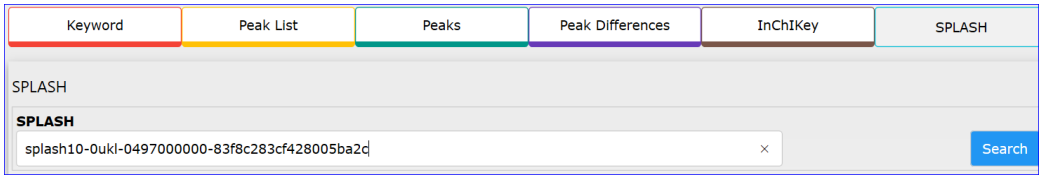


Figure 3.6.2. Lower panel for search target. All the mass spectra in MassBank are specified as SPLASH search target.

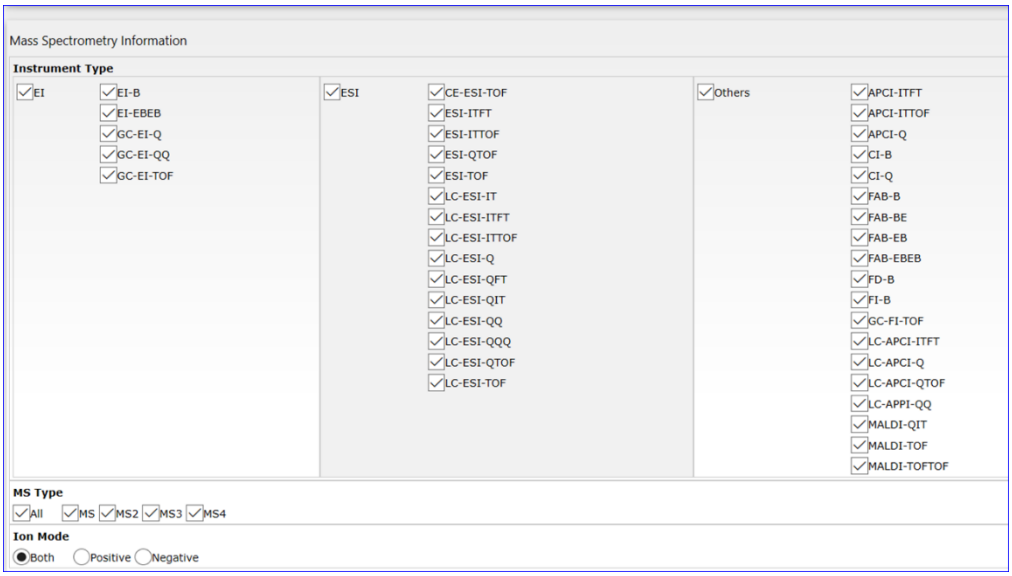
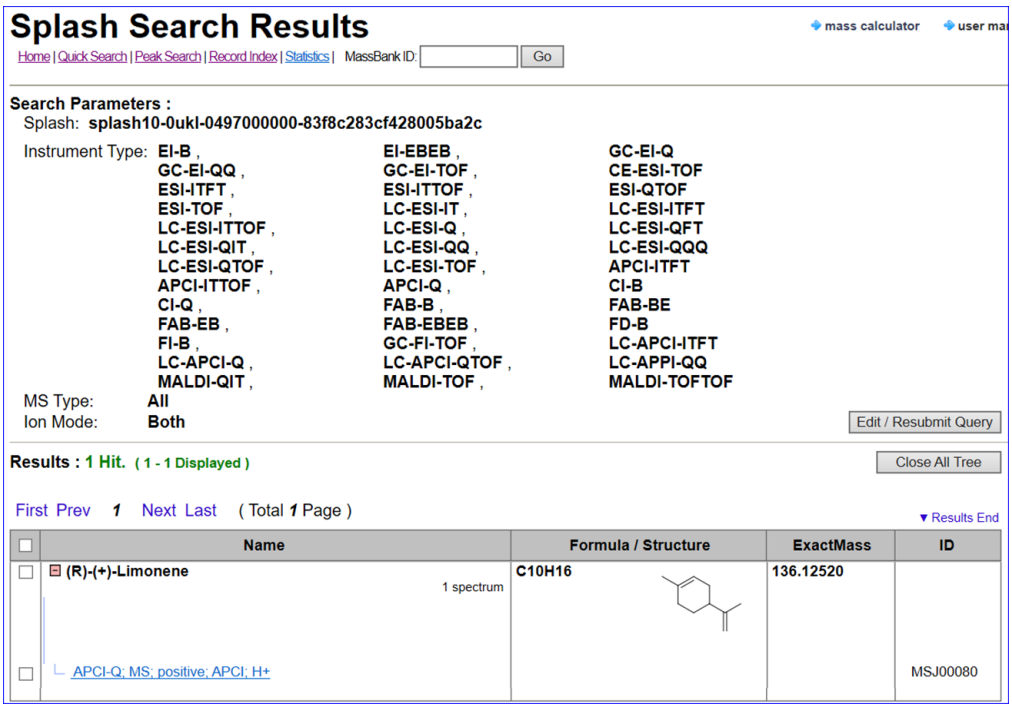


Figure 3.6.3. Summary of "SPLASH" search results on "Splash Search Results" panel.



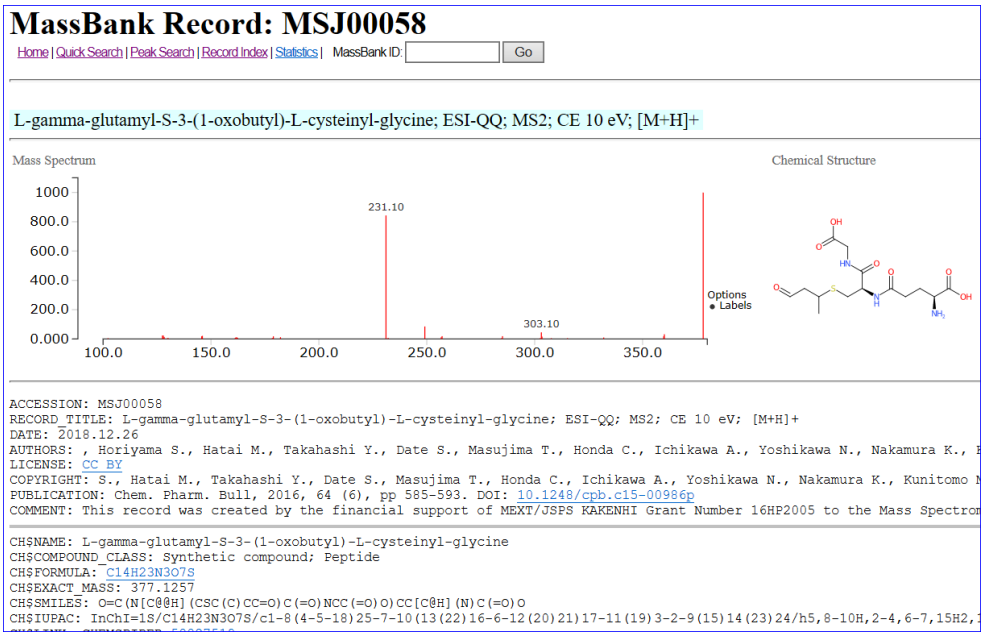
3.7. "MassBank Record ID" search

Function: This tool finds the mass spectrum specified by MassBank Record ID.

[1] Input MassBank Record ID in "MassBank ID | Go" or "MassBank ID | Go" that is displayed on the upper part of every page. For an example, see Figure 3.6.3.

[2] Crick "Go" button (It is not required to define search target).

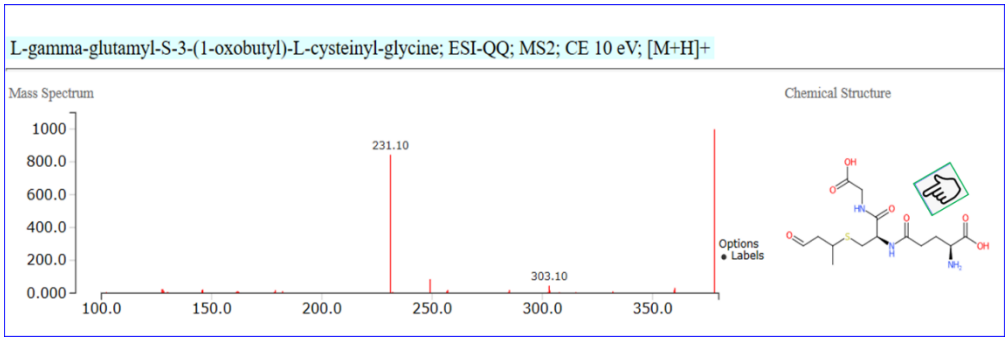
Figure 3.7.1. “MassBank Record ID” search displays “MSJ00058” record. Part of the record is shown.



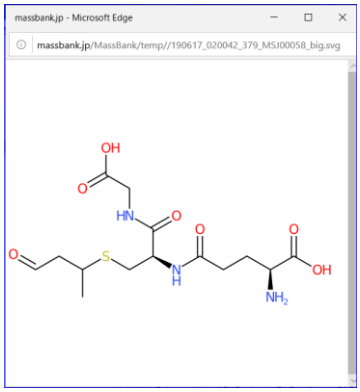
Tips: MassBank Record ID is “Accession number” correctly. It consists of 8 characters; the first 3 or 2 characters, Prefix of Accession number, define contributor group and the following five or six characters are record number within the contributor group. Current Prefix and contributors are listed in (<https://github.com/MassBank/MassBank-data/blob/master/List_of_Contributors_Prefixes_and_Projects.md>).

Tips： Zoom-in on the chemical structure.

Place a mouse cursor on the chemical structure and crick.



Enlarged structure displayed on another window. Crick “X” in the top right corner to close the window.

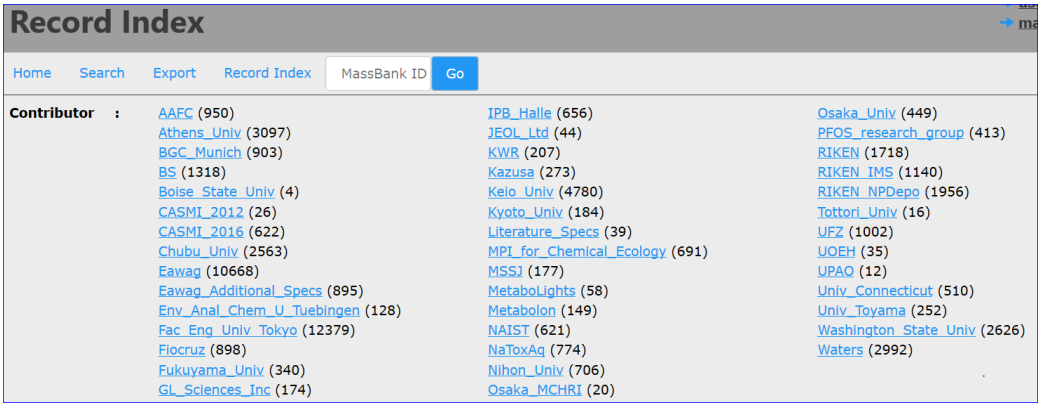


4. “Statistics of MassBank Data” page.

Function: This page is the summary of MassBank data statistics.

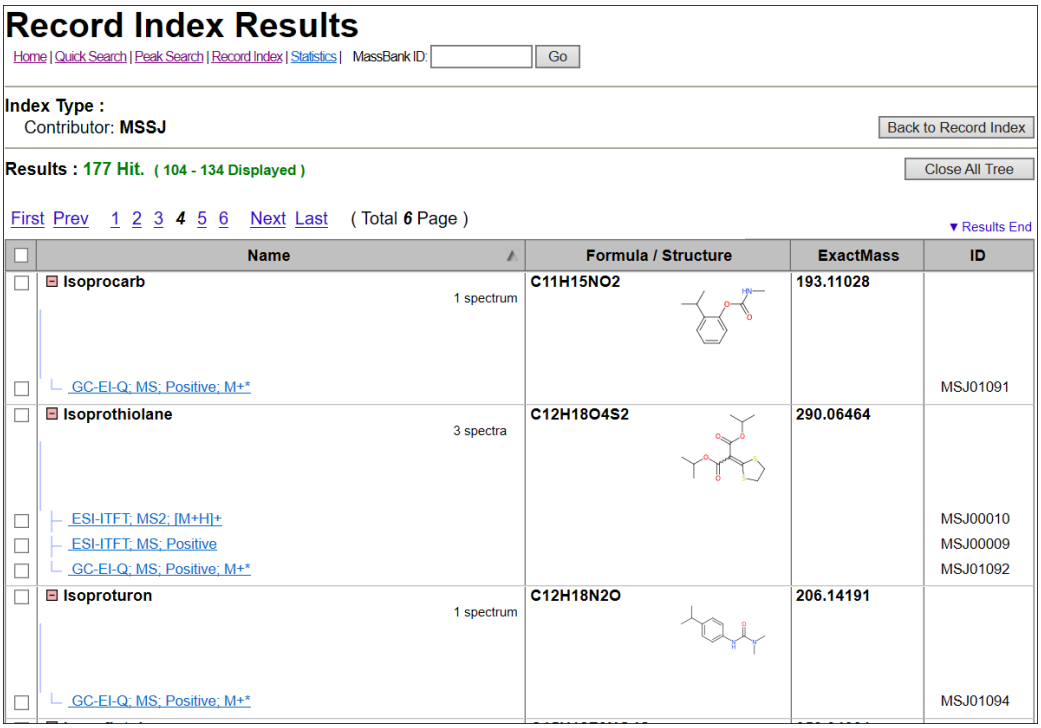
[1] Click Record Index on Figure 2 (page 2) to display Figures 4.1 and 4.3.

Figure 4.1. Record Index page consists of five sections. (1) “Contributor” section.



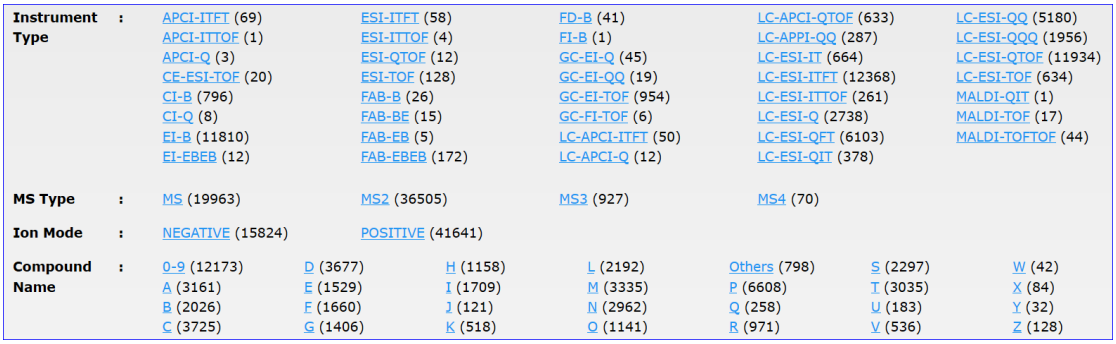
“Contributor” section summarizes the research groups that deposited their mass spectral data to MassBank. Each research group is listed by group name abbreviation followed by the number of mass spectral records in parentheses. Crick “group name abbreviation” to show its contribution (Figure 4.2).

Figure 4. 2. “Record Index Results” panel shows that the Mass Spectrometry Society of Japan (group name abbreviation = “MSSJ”) contributes a total of 177 mass spectral records.



Note 4.1: Mass spectrum analyzed by SIMS (MSJ00083) is not displayed. We will investigate the cause and fix it.

Figure 4. 3. Record Index page consists of five sections. (2) “Instrument Type”, (3) “MS Type”, (4) “Ion Mode”, and (5) “Compound Name” sections. The number in parentheses in each section is a total number of mass spectral records.



Instrument Type :	APCI-ITFT (69)	ESI-ITFT (58)	FD-B (41)	LC-APCI-QTOF (633)	LC-ESI-QQ (5180)		
	APCI-ITTOF (1)	ESI-ITTOF (4)	FI-B (1)	LC-APPI-QQ (287)	LC-ESI-QQQ (1956)		
	APCI-Q (3)	ESI-QTOF (12)	GC-EI-Q (45)	LC-ESI-IT (664)	LC-ESI-QTOF (11934)		
	CE-ESI-TOF (20)	ESI-TOF (128)	GC-EI-QQ (19)	LC-ESI-ITFT (12368)	LC-ESI-TOF (634)		
	CI-B (796)	FAB-B (26)	GC-EI-TOF (954)	LC-ESI-ITTOF (261)	MALDI-QIT (1)		
	CI-Q (8)	FAB-BE (15)	GC-FI-TOF (6)	LC-ESI-Q (2738)	MALDI-TOF (17)		
	EI-B (11810)	FAB-EB (5)	LC-APCI-ITFT (50)	LC-ESI-QFT (6103)	MALDI-TOFTOF (44)		
	EI-EBEB (12)	FAB-EBEB (172)	LC-APCI-Q (12)	LC-ESI-QIT (378)			
MS Type :	MS (19963)	MS2 (36505)	MS3 (927)	MS4 (70)			
Ion Mode :	NEGATIVE (15824)	POSITIVE (41641)					
Compound Name :	0-9 (12173)	D (3677)	H (1158)	L (2192)	Others (798)	S (2297)	W (42)
	A (3161)	E (1529)	I (1709)	M (3335)	P (6608)	T (3035)	X (84)
	B (2026)	F (1660)	J (121)	N (2962)	Q (258)	U (183)	Y (32)
	C (3725)	G (1406)	K (518)	O (1141)	R (971)	V (536)	Z (128)

If you have any problem on this manual, please contact us by email: massbank@mssj.jp.