



CEU MASS MEDIATOR USER'S MANUAL

Version 2.0, 31st July 2017



1. Introduction.....	2
1.1. System Requirements.....	2
2. Peak search.....	3
2.1. Simple Search	3
2.2. Advanced Search	5
2.3. Batch Search	7
2.4. Batch Advanced Search	8
2.5. Annotations rules	11
2.6. Submit menu	12
2.7. Result List	12
3. Pathway Displayer	0
3.1. File structure.....	0
3.2. Result list for pathways	1
4. Manual.....	2

1. Introduction

Ceu Mass Mediator (CMM) is an on-line tool for aiding researchers when performing metabolite annotation. CMM integrates compounds from different sources (HMDB, LipidMaps, KEGG and Agilent Metlin PCDL) based on the IUPAC International Chemical Identifier (InChI). Furthermore, CMM scores the putative annotations using three types of rules, explained in detail in section 2.5

This manual describes the available features in CMM. These features are shown in Figure 1 and described in chapter 2 and 3.

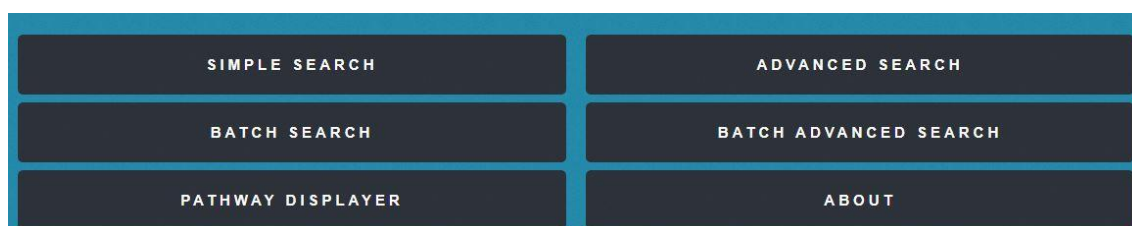


Figure 1 Main menu of Ceu Mass Mediator

1.1. System Requirements

CMM is a J2EE application, and it may be accessed through any web browser which supports JavaScript. CMM does not use Adobe Flash player neither popups. It has been tested in the next browsers:

- i. Mozilla Firefox 50
- ii. Google Chrome 45
- iii. Internet Explorer 11
- iv. Opera 42

2. Peak search

Peak Search allows the user to find metabolites based on the neutral or the m/z mass within a certain tolerance (default tolerance: 10 ppm). CMM enables 4 types of peak searches: simple and advanced search for single experimental mass: batch and batch advanced search for a set of experimental masses.

2.1. Simple Search

Simple search enables the user to find metabolites through the m/z or the neutral mass. Query parameters are specified in the form shown in Figure 2.

The figure shows a web-based search interface for metabolites. It consists of several labeled input fields and checkboxes:

- Experimental Mass:** A text input field with the placeholder "enter input mass" and a label [1].
- Tolerance (ppm):** A text input field with the value "10" and a label [2].
- Databases:** A list of checkboxes with labels: "All except MINE" (checked), "All (Including In Silico Compounds)", "Kegg", "HMDB", "LipidMaps", "Metlin", and "MINE (Only In Silico Compounds)". It has a label [3].
- Metabolites:** A list of radio buttons with labels: "All except peptides" (selected), "Only lipids", and "All including peptides". It has a label [4].
- Input Mass Mode:** A list of radio buttons with labels: "Neutral Masses" (selected) and "m/z Masses". It has a label [5].
- Ionization Mode:** A list of radio buttons with labels: "neutral" (selected), "Positive Mode", and "Negative Mode". It has a label [6].
- Adducts:** A list of checkboxes with labels: "All" (checked) and "M". It has a label [7].

Figure 2 Simple search interface

[1] **Experimental Mass (EM):** Mass to search in CMM (Da).

[2] **Tolerance:** Tolerance allowed for the putative annotations regarding the EM (ppm).

[3] **Databases:** The putative annotations should be present in the databases chosen by the user (Kegg, HMDB, LipidMaps, Metlin or MINE).

[4] **Metabolites:** Metabolite types to search. The user can filter the results based on the metabolite type. It may be used for excluding peptides, look only into lipids or perform a query over all type of metabolites.

[5] **Masses mode:** The user introduces the EM in neutral or m/z mode. If the user is working with neutral masses, CMM performs searches over positive or negative mode based on the hypothesis of the neutral mass calculated as $[M-H]^-$ or $[M+H]^+$. That means that the EM

will correspond to the m/z obtained in the mass spectrometer with the addition or subtraction of the mass of the hydrogen (H).

[6] Ionization mode: The user wants to perform searches over a mass obtained in positive or negative mode. Depending on the ionization mode, the possible adducts formed differ.

[7] Adducts: The possible adducts formed when running the experiment. The user may choose between different adducts in negative or positive mode. The list of possible adducts in negative and positive modes are shown in Figure 3 and Figure 4. All the possible alterations of the mass of the original metabolite (M) given by the selected adducts will be searched for by CMM.

The screenshot shows a software interface with two main sections. On the left, under the heading "Ionization Mode:", there is a list of three options: "neutral", "Positive Mode", and "Negative Mode". The "Negative Mode" option is highlighted with a dark background. Below this list, a text label reads "calculation of new m/z from neutral mass based on selected adducts". On the right, under the heading "Adducts:", there is a list of seven options, each preceded by a checkbox. The first option, "All", has its checkbox checked. The other options are "M-H", "M+Cl", "M+HCOO", "M-H-H2O", "M-H+HCOONa", and "2M-H".

Figure 3 Adducts to search in negative mode

The screenshot shows a software interface similar to Figure 3 but for positive ionization mode. Under the heading "Ionization Mode:", the "Positive Mode" option is highlighted. The text label below it remains "calculation of new m/z from neutral mass based on selected adducts". Under the heading "Adducts:", the "All" option is checked. The other options are "M+H", "M+Na", "M+K", "M+NH4", "M+H-H2O", "M+H+HCOONa", "M+2H", "2M+H", "2M+Na", and "2M+H-H2O".

Figure 4 Adducts to search in positive mode.

The only type of knowledge that may be applied in simple search corresponds to the ionization rules. Depending on the metabolite type, some adducts are expected to be formed, some others are possibly present and some others are not expected to appear. For more information, look into section 2.5.

2.2. Advanced Search

Advanced search enables the user to find metabolites through the m/z or the neutral mass including some extra query parameters that are not available in the simple search. In this section all the parameters are explained (see Figure 5).

The figure displays the 'Advanced Search' interface with the following components:

- Experimental Mass (*):** [1] Input field with placeholder 'enter input mass'.
- Tolerance -ppm- (*):** [2] Input field with value '10'.
- Retention Time:** [3] Input field with placeholder 'enter retention time'.
- Composite Spectrum:** [4] Input field with placeholder 'enter Composite Spectrum'.
- Chemical Alphabet (*):** [5] Dropdown menu with options: All, CHNOPS, CHNOPS + Cl.
- Modifiers (*):** [6] Dropdown menu with options: None, NH3, HCOO, CH3COO, HCOONH3, CH3COONH3.
- Databases (*):** [7] Checkable list:
 - ☒ All except MINE
 - ☐ All (Including In Silico Compounds)
 - ☐ Kegg
 - ☐ HMDB
 - ☐ LipidMaps
 - ☐ Metlin
 - ☐ MINE (Only In Silico Compounds)
- Metabolites (*):** [8] Dropdown menu with options: All except peptides, Only lipids, All including peptides.
- Input Mass Mode (*):** [9] Dropdown menu with options: Neutral Masses, m/z Masses.
- Ionization Mode (*):** [10] Dropdown menu with options: neutral, Positive Mode, Negative Mode.
- Adducts (*):** [11] Checkable list:
 - ☒ All
 - ☐ M

Figure 5 Advanced search interface

[1] **Experimental Mass (EM):** Mass to search in CMM (Da).

[2] **Tolerance:** Tolerance allowed for the putative annotations regarding the EM (ppm).

[3] Retention Time (RT): Amount of time spent by a compound on the column after it has been injected. It is an integer or a real number. The units used do not matter since it is used for checking relations between different putative annotations.

[4] Composite Spectrum (CS): Spectrum created by summation of all co-eluting m/z ions that are related, including isotopes, adducts and dimmers. It is used by CMM to calculate relations between them and automatically find which adduct corresponds to the peak, when more than one adduct is present in the CS; i.e., it is used to calculate which is the original mass whose alterations have given rise to the observed CS.

[5] Chemical Alphabet: Possible elements of the putative annotations. CHNOPS (carbon, hydrogen, nitrogen, oxygen, phosphorus, sulphur), CHNOPS + Cl (chlorum), all elements.

[6] Modifiers: Mobile phase modifier used. Depending on this modifier, the adduct formation may change. They are taken into account in the adduct formation rules (see section 2.5).

[7] Databases: The putative annotations should be present in the databases chosen by the user (Kegg, HMDB, LipidMaps, Metlin and/or MINE).

[8] Metabolites: Metabolite types to search. The user can filter the results based on the metabolite type. It may be used for excluding peptides, look only into lipids or perform a query over all type of metabolites.

[9] Masses mode: The user introduces the EM in neutral or m/z mode. If the user is working with neutral masses, CMM performs searches over positive or negative mode based on the hypothesis of the neutral mass calculated as $[M-H]^-$ or $[M+H]^+$. That means that the EM will correspond to the m/z obtained in the mass spectrometer with the addition or subtraction of the mass of the hydrogen (H).

[10] Ionization mode: The user wants to perform searches over a mass obtained in positive or negative mode. Depending on the ionization mode, the possible adducts formed differ.

[11] Adducts: The possible adducts formed when running the experiment. The user may choose between different adducts in negative or positive mode. The list of possible adducts in negative and positive modes are shown in Figure 3 and Figure 4. All the possible alterations of the mass of the original metabolite (M) given by the selected adducts will be searched for by CMM.

The knowledge that may be applied in advanced search corresponds to the ionization rules. Depending on the metabolite type, some adducts are expected to be formed, some others are possibly present and some others are not expected to appear. For more information, look into section 2.5. However, the rules about adduct formation and lipid elution time cannot be applied since they are based in the relations between different peaks, and advanced search only accepts one peak.

2.3. Batch Search

Batch search enables the user to find metabolites through the m/z or the neutral masses. Query parameters are specified in the form shown in Figure 6.

The screenshot displays the Batch Search interface with the following components:

- Experimental Masses:** A text input field labeled "enter input masses" with a [1] annotation.
- Tolerance (ppm):** A numeric input field set to "10" with a [2] annotation.
- Databases:** A list of checkboxes for database selection: "All except MINE" (checked), "All (Including In Silico Compounds)", "Kegg", "HMDB", "LipidMaps", "Metlin", and "MINE (Only In Silico Compounds)" with a [3] annotation.
- Metabolites:** A dropdown menu with options "All except peptides", "Only lipids", and "All including peptides" with a [4] annotation.
- Input Masses Mode:** A dropdown menu with "Neutral Masses" (selected) and "m/z Masses" with a [5] annotation.
- Ionization Mode:** A dropdown menu with "neutral" (selected), "Positive Mode", and "Negative Mode" with a [6] annotation.
- Adducts:** A dropdown menu with "All" (checked) and "M" with a [7] annotation.

Figure 6 Batch search interface

[1] Experimental Masses (EM): Masses to search in CMM (Da).

[2] Tolerance: Tolerance allowed for the putative annotations regarding the EM (ppm).

[3] Databases: The putative annotations should be present in the databases chosen by the user (Kegg, HMDB, LipidMaps, Metlin or MINE).

[4] Metabolites: Metabolite types to search. The user can filter the results based on the metabolite type. It may be used for excluding peptides, look only into lipids or perform a query over all type of metabolites.

[5] Masses mode: The user introduces the EM in neutral or m/z mode. If the user is working with neutral masses, CMM performs searches over positive or negative mode based on the hypothesis of the neutral mass calculated as $[M-H]^-$ or $[M+H]^+$. That means that the EM will correspond to the m/z obtained in the mass spectrometer with the addition or subtraction of the mass of the hydrogen (H).

[6] Ionization mode: The user wants to perform searches over a mass obtained in positive or negative mode. Depending on the ionization mode, the possible adducts formed differ.

[7] Adducts: The possible adducts formed when running the experiment. The user may choose between different adducts in negative or positive mode. The list of possible adducts in negative and positive modes are shown in Figure 3 and Figure 4. All the possible alterations of the mass of the original metabolite (M) given by the selected adducts will be searched for by CMM.

2.4. Batch Advanced Search

Batch advanced search enables the user to find metabolites through the m/z or the neutral masses query parameters explained in section 2.2. In addition, it has three input fields devoted to biomarker discovery experiments. The experimental masses corresponding to non-significant features together with its corresponding RT and CS may be introduced to provide evidences that support or refute the putative annotations. However, the putative annotations of the compounds introduced in all experimental masses field, but not included in significant experimental masses, are not returned in the result list.

Figure 7 shows the fields of the batch advanced search. The only mandatory field regarding to the features obtained in the mass spectrometer are the experimental masses of the significant compounds. RT, CS and non-significant experimental masses are optional fields that will be used by CMM for applying knowledge based on the rules explained in section 2.5. The more information the user provides in the form, the more evidence can be used for supporting or refuting the putative annotations.

[1] Significant experimental Masses (EM): Masses to search in CMM (Da) corresponding to significant features extracted after the statistical analysis.

[2] Retention Time (RT): Amount of time spent by a compound on the column after it has been injected. It is an integer or a real number. The units used do not matter since it is used for checking relations between different putative annotations. The RT here introduced correspond to the experimental masses introduced in field **[1]** in the order they are introduced.

[3] Composite Spectrum (CS): Spectrum created by summation of all co-eluting m/z ions that are related, including isotopes, adducts and dimmers. It is used by CMM to calculate relations between them and automatically find which adduct corresponds to the peak (when more than one adduct is present in the CS). The CS here introduced correspond to the experimental masses introduced in field **[1]** in the order they are introduced.

[4] All experimental Masses (EM): All masses corresponding to significant and non-significant features extracted after the statistical analysis. Non-significant masses provide evidence for supporting or refuting the putative annotations, but are not returned among the results of the query.

[5] All Retention Times (RT): Amount of time spent by a compound on the column after it has been injected. Unity used does not really matter since it is used for checking

relations between different putative annotations. The RTs here introduced correspond to the experimental masses introduced in field **[4]** in the order they are introduced.

Experimental Masses (*): [1] enter input masses	Retention Times: [2] enter Retention Times	Composite Spectrums: [3] enter Composite Spectrum
All Experimental Masses: [4] enter all input masses	All Retention Times: [5] enter Retention Times	All Composite Spectrums: [6] enter Composite Spectrum
Tolerance -ppm- (*): 10 [7]		
Chemical Alphabet (*): All [8] CHNOPS CHNOPS + Cl		
Modifiers (*): None [9] NH3 HCOO CH3COO HCOONH3 CH3COONH3		
Databases (*): <input checked="" type="checkbox"/> All except MINE [10] <input type="checkbox"/> All (Including In Silico Compounds) <input type="checkbox"/> Kegg <input type="checkbox"/> HMDB <input type="checkbox"/> LipidMaps <input type="checkbox"/> Metlin <input type="checkbox"/> MINE (Only In Silico Compounds)		
Metabolites (*): All except peptides [11] Only lipids All including peptides		
Input Masses Mode (*): [12] Neutral Masses m/z Masses	Ionization Mode (*): [13] neutral Positive Mode Negative Mode	Adducts (*): [14] <input checked="" type="checkbox"/> All <input type="checkbox"/> M

Figure 7 Batch advanced search interface

[6] All Composite Spectrum (CS): Spectrum created by summation of all co-eluting m/z ions that are related, including isotopes, adducts and dimmers. It is used by CMM to calculate relations between them and automatically find which adduct corresponds to the peak (when more than one adduct is present in the CS) i.e., it is used to calculate which is the original mass whose alterations have given rise to the observed CS. The CSs introduced here correspond to the experimental masses introduced in field **[4]** in the order they are introduced.

[7] Tolerance: Tolerance allowed for the putative annotations regarding the significant EM (ppm).

[8] Chemical Alphabet: Possible elements of the putative annotations. CHNOPS, CHNOPS + Cl, all elements.

[9] Modifiers: Mobile phase modifier used. Depending on this modifier, the adduct formation may change. They are taken into account in the adduct formation rules (see section 2.5).

[10] Databases: The putative annotations should be present in the databases chosen by the user (Kegg, HMDB, LipidMaps, Metlin and/or MINE).

[11] Metabolites: Metabolite types to search. The user can filter the results based on the metabolite type. It may be used for excluding peptides, look only into lipids or perform a query over all type of metabolites. CMM considers as lipids the compounds present in LipidMaps.

[12] Masses mode: The user introduces the EM in neutral or m/z mode. If the user is working with neutral masses, CMM performs searches over positive or negative mode based on the hypothesis of the neutral mass calculated as $[M-H]^-$ or $[M+H]^+$. That means that the EM will correspond to the m/z obtained in the mass spectrometer with the addition or subtraction of the mass of the hydrogen (H).

[13] Ionization mode: The user wants to perform searches over a mass obtained in positive or negative mode. Depending on the ionization mode, the possible adducts formed differ.

[14] Adducts: The possible adducts formed when running the experiment. The user may choose between different adducts in negative or positive mode. The list of possible adducts in negative and positive modes are shown in Figure 3 and Figure 4. All the possible alterations of the mass of the original metabolite (M) given by the selected adducts will be searched for by CMM.

Batch advanced search process all information provided (significant EM are mandatory, RT, CS and non-significant EM are optional) for scoring the putative annotations based on the rules explained in section 2.5

2.5. Annotations rules

Ceu Mass Mediator scores the putative annotations based on expert knowledge. This knowledge applied is especially devoted to lipids using Liquid Chromatography. It uses 143 rules divided in three main types:

1. Propensity of particular adducts formation depending on the lipid class, ionisation mode and mobile phase modifier used. Lipids belonging to particular class may always form some adducts in certain experimental conditions, whereas they may form others in different conditions. The mobile phase modifier used is indicated manually by the user. For example, phosphocholine in negative mode primarily form $[M+HCOO]^-$ or $[M+CH_3COO]^-$ depending on the modifier used ($HCOO^-$ or CH_3COO^-); they may also form $M+Cl^-$ with lower intensity; and they never form $M-H^-$. Lipid classes used in these rules are: PC, LPC, PE, LPE, PI, PG, PS, LPS, PA, MG, DG, TG, CER, SM and CE according to the LipidMaps classification.
2. Relationship between signals of different adducts from the same compound (Lynn et al., 2015). We only expect certain types of adducts when others are present. For example, glycerophosphoethanolamines (PE) may form $M+Na^+$ adduct, but only when $M+H$ adduct is also formed in higher abundance. If an experimental mass (738.5044 Da) is compatible with a $M+Na^+$ adduct of PE(34:2), but the adduct $M+H^+$ (716.5225 Da) is not present in the whole data matrix, CMM decreases the score of the annotation of PE(34:2) for experimental mass 738.5044 Da and adduct $M+Na^+$.
3. Relative RT based on the lipid class and the length and number of double bounds in the lipid carbon chains (Godzien et al., 2016). For example, RT of LPG(18:0) must be greater than RT of LPG(16:0); and RT of LPG(18:0) must be greater than RT of LPG(18:2).

CMM calculates a score for each of these three rule types (χ_1 , χ_2 , χ_3) and then it integrates them by computing their weighted geometric mean:

$$\chi = \exp\left(\frac{\sum_{i=1}^3 \omega_i \cdot \ln \chi_i}{\sum_{i=1}^3 \omega_i}\right)$$

where ω_i is the weight of each score and χ_i is the punctuation for each score. $\omega_1 = 1$, $\omega_2 = 1$ and $\omega_3 \in [0, 2]$. ω_3 depends on the number of rules applied for lipid elution time. This is the only rule type that can be triggered a variable number of times for the same annotation, depending on how many other lipid annotations with which the retention time of the annotation to be scored can be compared with. The more rules have been triggered, the more evidence supporting or refuting the annotation would have been gathered, the more weight this evidence should have on the final score. Internally all $\chi_i \in [0, 1]$, corresponding 0 with a completely refuted annotation, 1 with an annotation for which all the possible evidence is available and it is positive, and the value of 0.5 with an annotation for which there is no evidence (neither refuting nor supporting) but the annotation's mass matches the query parameters. However, scores are multiplied by 2 in the user interface because our experience

has shown us that it is more intuitive to the researchers to see a final score in the interval [0, 2].

2.6. Submit menu

Once the user has performed any type of query explained in sections 2.1, 2.2, 2.3, and 2.4, the query is sent to the server when the button submit compounds (See [2] of Figure 8)

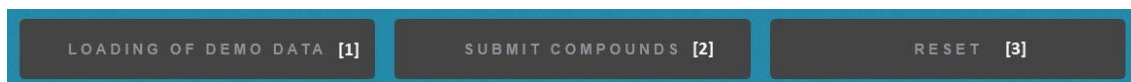


Figure 8 Submit compounds menu

[1] **Loading demo data:** Demo data is loaded. User data is lost.

[2] **Submit compounds:** Submit query with the filled fields by the user.

[3] **Reset:** Clears the fields to start again filling the query parameters and input fields.

2.7. Result List

Once the user has performed any type of query explained in sections 2.1, 2.2, 2.3, and 2.4, a list of results is returned by CMM. Figure 9 shows an example of a result list.

[1] **Compound Id:** CMM Id.

[2] **Name:** Name of the putative annotation compound.

[3] **Formula:** Formula of the putative annotation compound.

[4] **Molecular weight:** Molecular weight of the putative annotation compound.

[5] **Retention time:** Retention time introduced by the user for the experimental mass (see [18]).

[6] **Error PPM:** Difference in parts per million (ppm) between the molecular weight and the corresponding experimental mass ([18]) and its corresponding adduct ([19]).

[7] **Score 1:** Score for ionization rules (see item 1 of section 2.5). The code colour is structured in four ranges.

[0, 0.5) is red and means that this annotation is very likely wrong.

[0.5, 1) is orange and means that this annotation is likely wrong.

[1, 1.5) is yellow and means that this annotation is likely right.

[1.5, 2] is green and means that this annotation is very likely right.

[8] **Score 2:** Score for adduct formation rules (see item 2 of section 2.5). The code colour is the same than for score 1 (see [7]).

[9] Score 3: Score for lipid elution order (see item 3 of section 2.5). The code colour is the same than for score 1 (see **[7]**).

[10] Final score: Integrated score (see section 2.5). The code colour is the same than for score 1 (see **[7]**).

[11] Cas: CAS Id.

[12] KEGG Id: KEGG ID and its corresponding link.

[13] HMDB Id: HMDB ID and its corresponding link.

[14] LipidMaps Id: LipidMaps ID and its corresponding link.

[15] Metlin Id: Metlin ID and its corresponding link.

[16] PubChem Id: Pub Chemical Id and its corresponding link.

[17] Pathways: Pathways from KEGG where the compound is present and its corresponding link.

[18] Experimental mass: Experimental mass introduced by the user.

[19] Adduct: Corresponding adduct for this table.

[20] Number of hits: Number of hits found for the search corresponding to experimental mass (**[18]**) and its corresponding adduct (**[19]**).

[21] Generate Excel: Button which generates an Excel file with the complete result list (all experimental masses and adducts). This excel file contains the same fields that the on-line interface, the same code colour explained in **[7]**.

GENERATE EXCEL

[21]

Results

[18]

[19]

[20]

11

12

13

14

15

16

17

18

19

20

Metabolites found for mass 495.3352 and adduct M+H -> 9 metabolites found

Id	Name	Formula	Molecular Weight	Retention Time	error PPM	Score1	Score2	Score3	Final Score	Cas	KEGG	HMDB	LipidMaps	Metlin	PubChem	Pathways
[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]	[13]	[14]	[15]	[16]	[17]
32773	PC(0.0/16.0)	C24H50NO7P	495.332492	19.46886	5	2.0	2.0	2.0	2.0				LMGP01050074	49340		
32785	PC(16.0/0.0)[rac]	C24H50NO7P	495.332492	19.46886	5	2.0	2.0	2.0	2.0				LMGP01050113	102768		
34165	PE(19.0/0.0)	C24H50NO7P	495.332492	19.46886	5	2.0	2.0	2.0	2.0				LMGP02050028	77694		
32409	PC(O-14.0/2.0)	C24H50NO7P	495.332492	19.46886	5	2.0	2.0	N/A	2.0				LMGP01020019	40048		
101416	PC(O-14.0/2.0)[U]	C24H50NO7P	495.33248947	19.46886	5	N/A	2.0	N/A	2.0					40049		
101417	PC(16.0/0.0)[S]	C24H50NO7P	495.33248947	19.46886	5	N/A	2.0	N/A	2.0					40285		
101418	PC(16.0/0.0)[U]	C24H50NO7P	495.33248947	19.46886	5	N/A	2.0	N/A	2.0					40286		
101419	PC(0.0/16.0)[U]	C24H50NO7P	495.33248947	19.46886	5	N/A	2.0	N/A	2.0					40341		
32744	PC(16.0/0.0)	C24H50NO7P	495.33248947	19.46886	5	2.0	2.0	2.0	2.0			HMDB10382	LMGP01050018	40284	460602	SHOW PATHWAYS

Metabolites found for mass 495.3352 and adduct M+Na -> 1 metabolites found

Id	Name	Formula	Molecular Weight	Retention Time	error PPM	Score1	Score2	Score3	Final Score	Cas	KEGG	HMDB	LipidMaps	Metlin	PubChem	Pathways
0	No compounds found for experimental mass 495.3352 and adduct: M+Na because we detected the adduct based on the composite spectrum. Look results for adduct: M+H		0.0	19.46886	0	N/A	N/A	N/A	N/A							

Metabolites found for mass 495.3352 and adduct M+K -> 1 metabolites found

Id	Name	Formula	Molecular Weight	Retention Time	error PPM	Score1	Score2	Score3	Final Score	Cas	KEGG	HMDB	LipidMaps	Metlin	PubChem	Pathways
0	No compounds found for experimental mass 495.3352 and adduct: M+K because we detected the adduct based on the composite spectrum. Look results for adduct: M+H		0.0	19.46886	0	N/A	N/A	N/A	N/A							

Figure 9 Result list interface

3. Pathway Displayer

This feature extract the information of a list of already identified compounds in order to perform a rank about the pathways that are more probably affected based on two different parameters: specificity of the compounds and percentage of compounds of the complete pathway from KEGG present in the file.

3.1. File structure

To upload an excel file to be analysed by pathway displayer of CMM, it is needed to press the button Choose file and, once the file was selected, submit it (see Figure 10). The structure of the file should follow the structure of the downloaded files from the result list (see Figure 11). The header names of lines 1 and two should be present in the file, and pathways are listed in subsequent columns after the column T.

The user should filter the result list until it only contains the annotations corresponding to the identified compounds. If the user has worked with CMM, these annotations have a list of pathways where the compound is present according to KEGG database.

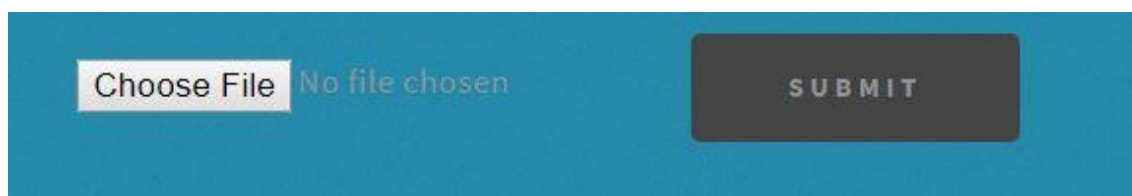


Figure 10 Pathway displayer menu

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	
1	LIST OF COMPOUNDS	Retention Time	Identifier	Adduct	PPM Error	Molecular Weight	Name	Formula	Score 1	Score 2	Score 3	Final Score	CAS	KEGG	HMDB	LipidMaps	Metlin	PubChem	InChIKey	Pathways		
2	Experimental mass	838.5571	27.7552	81516 M+H	216	838.3757	Heme O	C49H58FeN	N/A	N/A	N/A	N/A	13739	C1567	HMDB	6044	6044	F8P8ASV	Porphyrin Metab.	Metabolic pathways		
3		838.5571	27.7552	17602 M+H	647	839.1	oxalyl-CoA	C23H36O7	N/A	N/A	N/A	N/A		C0031	LMFA0705	52424	52424	OVXNZFT	Metabolic pathway: Glyoxylate and dicar			
4		838.5571	27.7552	17584 M+H	690	838.1363	Lacetyl-CoA	C24H40N7O	N/A	N/A	N/A	N/A		C0092	HMDB	LMFA0705	52424	52424	VWKEBO	Propanoate Metab	Microbial metabolism	
5		838.5571	27.7552	17471 M+H	690	838.1363	3-hydroxypropanoyl-	C24H40N7O	N/A	N/A	N/A	N/A		C0586	LMFA0705	440753	440753	BERBFZC	Propanoate Metab	Beta-Alanine Metabo		
6		838.5571	27.7552	17527 M+NH4	493	821.1258	acrylyl-CoA	C24H38N7O	N/A	N/A	N/A	N/A		C0089	LMFA0705	439349	439349	POODSGU	Propanoate Metab	Beta-Alanine Metabo		
7		838.5571	27.7552	96794 2M+H	60	419.3036	Myxalamid S	C25H41NO4	N/A	N/A	N/A	N/A		C1215				68320	11953999	OADGIMS-	Type I polyketide structures	
8		838.5571	27.7552	92133 2M+H	78	419.246	13-Desoxyxaxiline	C27H33NO3	N/A	N/A	N/A	N/A		C2053						SJZCKCB	Biosynthesis of	Indole diterpene alkaloids
9		838.5571	27.7552	91096 2M+H	338	419.1369	Jadomycin B aglycon	C24H21NO6	N/A	N/A	N/A	N/A		C1869				72460	12180163	HKQYQO	Biosynthesis of	Biosynthesis of ty
10		838.5571	27.7552	74977 2M+H	588	419.0321	S-(2,2-Dichloro-1,4,4'-Diapophytene	C12H19Cl2N	N/A	N/A	N/A	N/A		C1486	HMDB			76368	11954070	QVLCBMM	Metabolism of xenobiotics by cytochr	
11		838.5571	27.7552	88618 2M+Na	215	408.3756	15-cis-4,4'-Bile salt	C30H48	N/A	N/A	N/A	N/A		C1614				64111	14019219	IKGLUFES3	Carotenoid biosynthesis	
12		838.5571	27.7552	93969 2M+Na	0	408.2876	Bile acid	C24H40O5	N/A	N/A	N/A	N/A		C0156				438520		YRZLCOQ-	Vitamin digestion and ab	
13		838.5571	27.7552	2050 2M+Na	0	408.2876	Allocholic acid	C24H40O5	N/A	N/A	N/A	N/A		C1773	LMST0401			160638		YRZLCOQ-	Secondary bile acid biosynthesis	
14		838.5571	27.7552	2051 2M+Na	0	408.2876	Trihydroxy-Salpa-	C24H40O5	N/A	N/A	N/A	N/A		C1785	LMST0401			42696	5263852	PSHXEQJ	Secondary bile acid biosynthesis	
15		838.5571	27.7552	2059 2M+Na	0	408.2876	Haemuloholic acid	C24H40O5	N/A	N/A	N/A	N/A		C1786	LMST0401			42701	5263886	PPFQHK-	Secondary bile acid biosynthesis	
16		838.5571	27.7552	2060 2M+Na	0	408.2876	Phocaecholic acid	C24H40O5	N/A	N/A	N/A	N/A		C1785	LMST0401			42696	5263852	PPFQHK-	Secondary bile acid biosynthesis	
17		838.5571	27.7552	2065 2M+Na	0	408.2876	Avicholic acid	C24H40O5	N/A	N/A	N/A	N/A		C1786	LMST0401			42701	5263886	PPFQHK-	Secondary bile acid biosynthesis	
18		719.5465	27.7563	85253 M+H	140	719.4456	Erythromycin C	C36H65NO1	N/A	N/A	N/A	N/A		C0681				83833		RXPFWNT-	Biosynthesis of	Biosynthesis of 12-
19		719.5465	27.7563	2850 M+H	140	719.4456	Erythromycin C	C36H65NO1	N/A	N/A	N/A	N/A						LMPK0400		RXPFWNT-		
20		719.5465	27.7563	102793 M+H	140	719.4456	Erythromycin C	C36H65NO1	N/A	N/A	N/A	N/A	1675-					74575				

Figure 11 Structure of the Excel file for pathway displayer

Once the excel file is loaded, CMM processes it taking into account two different parameters for ordering the pathways present in the excel file. This order may guide the researcher to focus his hypothesis in these pathways that have compounds more specific (For example, Chlorophyll is only present in pathways related to plants):

1. Specificity: In how many pathways is present the compound? It uses the formula:

$$\text{Specificity} = \text{Min} \left(\frac{1}{\text{number of pathways where the compound has been detected}} \right)$$

Specificity $\in (0,1]$.

- Percentage of the compounds: How many compounds of the pathway are present? It uses the formula:

$$\text{Percentage} = \frac{\text{Number of compounds present in the file in pathway}}{\text{Total number of compounds present in the pathway}}$$

Percentage $\in (0,1]$.

The final order is determined by specificity and percentage. Specificity is the first parameter and, if the specificity is the same, then the percentage would be taken into account.

3.2. Result list for pathways

When the excel file is processed, CMM returns to the user a list of results with the pathways ordered (see section 3.1). Figure 12 shows an example of a list of pathways present in an excel file ordered using this approach. The results are also available in excel format if the user wants to work with it.

Compounds present in Quinolones											
Experimental mass	Retention Time	Id	Adduct	error PPM	Molecular Weight	Name	Formula	Cas	KEGG	HMDB	Lipid
719.5465	27.7563	94019	2M+Na	880	349.0896	Ulfloxacin; NM 394	C16H16FN3O3S	112984-60-8	C14492		
< >											
Compounds present in Opioid receptor agonists/antagonists											
Experimental mass	Retention Time	Id	Adduct	error PPM	Molecular Weight	Name	Formula	Cas	KEGG	HMDB	Lipid
750.5411	27.7562	94531	M+H-H2O	222	768.3806	Deltorphan C; Deltorphan I	C37H52N8O10	122752-15-2	C18097		
< >											
Compounds present in Aflatoxin biosynthesis											
Experimental mass	Retention Time	Id	Adduct	error PPM	Molecular Weight	Name	Formula	Cas	KEGG	HMDB	Lipid
676.5043	5.7029	92777	2M+H	620	338.0427	Versicolorin A	C18H10O7	6807-96-1	C20583		
719.5465	27.7563	88755	2M+H	864	360.0845	Versiconol	C18H16O8	22268-13-9	C20508		
750.5411	27.7562	92616	2M+H-H2O	498	384.0845	1'-Hydroxyversicolorone; Hydroxyversicolorone	C20H16O8	111975-78-1	C20503		
761.5935	27.7564	92948	2M+Na	809	370.1053	Norsolorinic acid; Norsolorinate; 2-Hexanoyl-1,3,6,8-tetrahydroxy-9,10-anthraquinone	C20H18O7	10254-99-6	C20452		

Figure 12 Results list of the pathway displayer

4. Manual

This section corresponds to the download of different versions of the manuals in PDF.

In user's manual page, a list of available user's manuals is presented. Nowadays, only the user's manual version 2.0 is available (see Figure 13), but it will be updated as soon as new features will be available in the CMM.

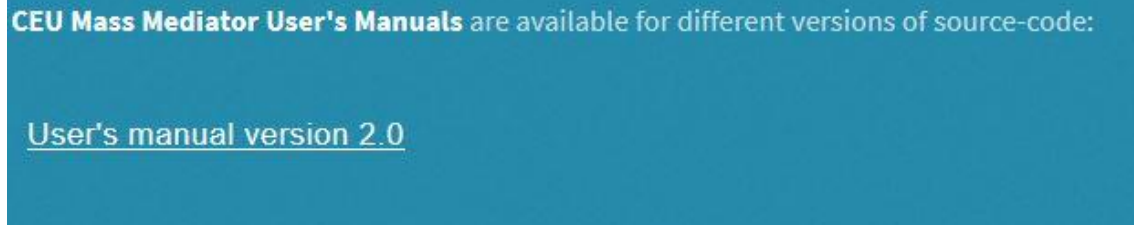


Figure 13 User's manual page