# A Carbonyl-Trapping Mechanism-based Automatic Mining (CTM-AM) Strategy for Accelerating Discovery of Natural Products with Anti-Advanced Glycation End Products Activity

## User Manual of CTM-AM

Wenjun Qi, Mi Zhang, Takashi Kikuchi*, Kouharu Otsuki, Wei Li*

Faculty of Pharmaceutical Sciences, Toho University, Miyama 2-2-1, Funabashi, Chiba 274-8510, Japan

E-mail: liwei@phar.toho-u.ac.jp

## Software Introduction

**CTM-AM** was employed to accelerate the discovery of anti-AGE natural products. This method facilitates the construction of a methylglyoxal (MGO) adduct library, guided by carbonyl trap reaction properties, for the rapid identification of MGO-trapped components in medicinal plants. Moreover, the tool offers a flexible framework that can be readily adapted to other reactive dicarbonyl species.

## Software Setup

Download and install R with R studio (R 3.6.0 or higher to be installed).

1. Required Packages:

"Shiny, DT, dplyr, fs, purrr, readr, readxl, writexl, stringr, shinyWidgets, bslib"

2. Launching the Application:

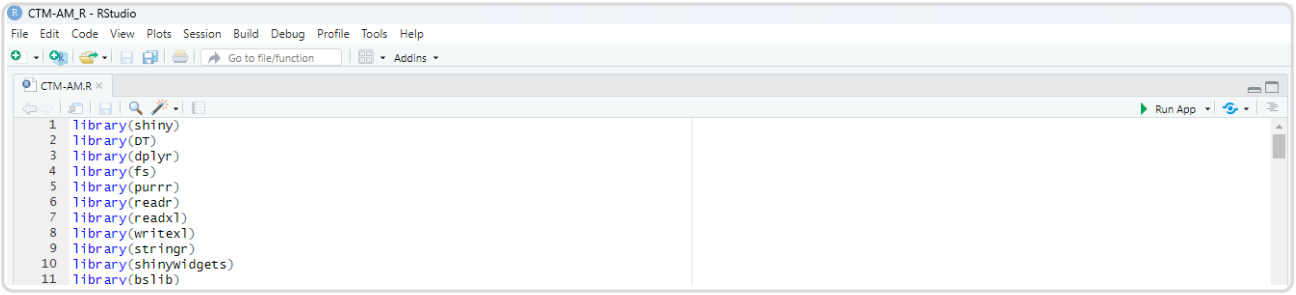


**Fig 1.** Enter the main interface of CTM-AM.

(1) Open RStudio.

(2) Open the application's main R script "CTM-AM.R".

(3) Click the "Run App" button in the top-right corner of the script window.

(4) The application will open in default web browser or inside RStudio's Viewer panel.

# Graphical User Interface

## 1. Construction of MGO Adducts Library

The "Tab-Library" allows users to generate a corresponding molecular weight library by adjusting the number of attachable carbonyl compounds.

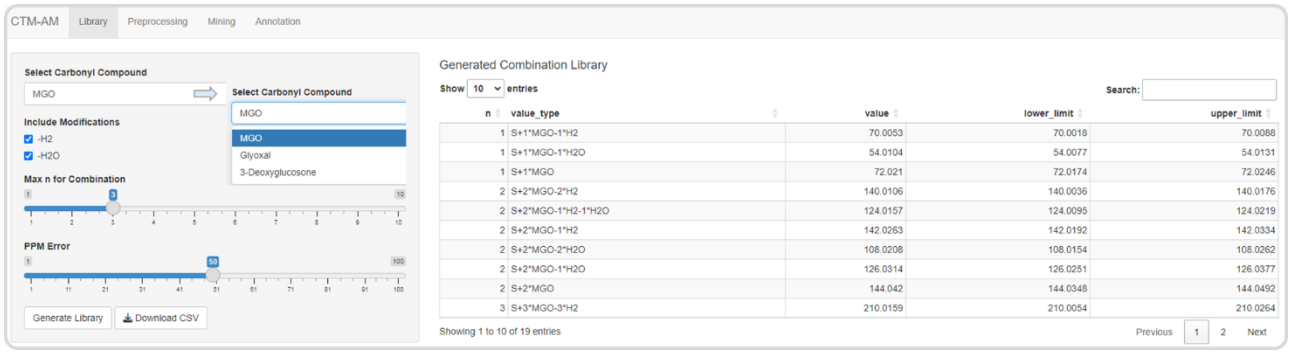


**Fig 2.** The Library function of the CTM-AM interface.

(1) Select the type of carbonyl compound and reaction product and determine the number of reactants that can be reacted.

(2) Set molecular weight range (ppm).

(3) Click the "Generate Library" button to view the table and save it in csv file format.

## 2. Align data preprocessing and classification

The "Tab-Preprocessing" module allows users to group and classify files processed by MZmine for subsequent feature matching and filtering.

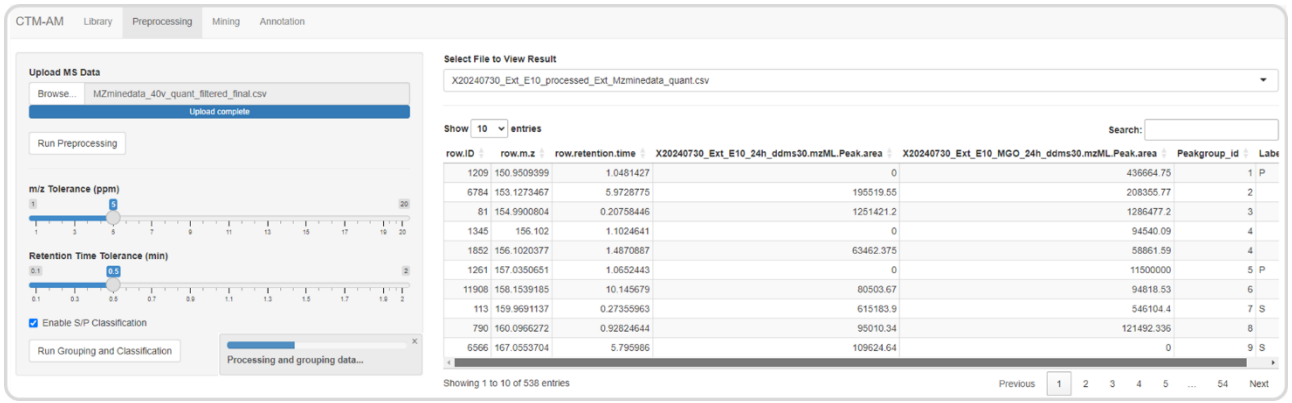


**Fig 3.** The Preprocessing function of the CTM-AM interface.

(1) Import an alignment file in .csv format and match the data before and after the same extraction reaction.

(2) Set the matching error of peaks based on m/z (ppm) and retention time (min).

(3) Click the "Run Grouping and Classification " button.

### 3. Product library MS1 matching and MS2 spectrum mining

The "Tab-Mining " automatically identifies compounds and their corresponding products at both MS1 and MS2 levels, guided by the characteristics of carbonyl trap reactions.

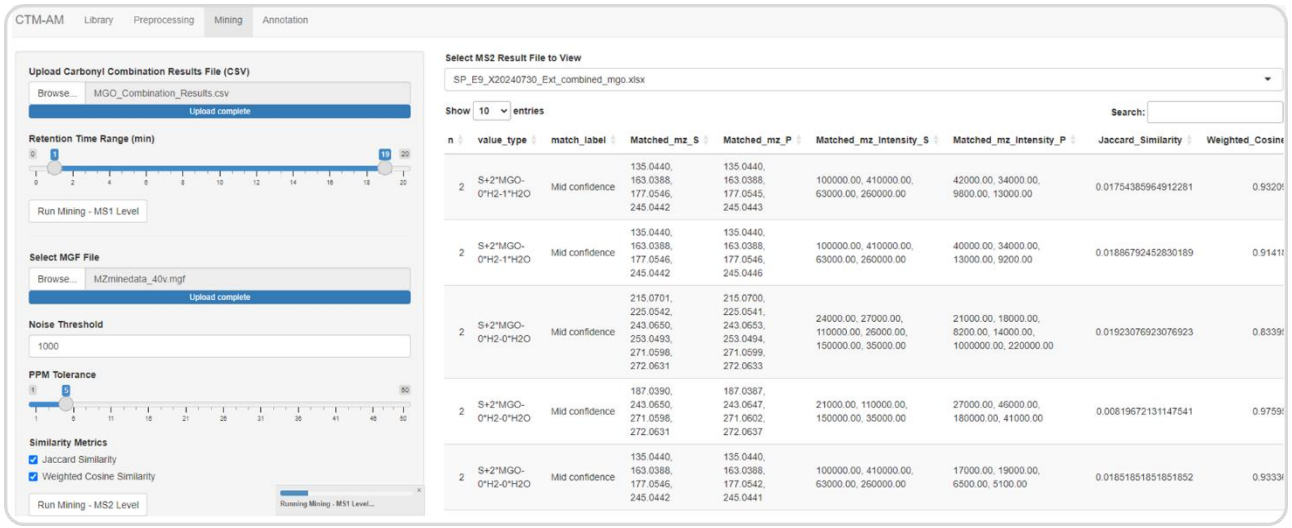


**Fig 4.** The Mining function of the CTM-AM interface.

(1) Select the .csv file of the generated carbonyl product library and explore the retention time range of active compounds.

(2) Click the "Run Mining-MS1 Level" button to perform matching mining.

(3) Select the .mgf file of MZmine processed extract data. Set the noise threshold and error for MS2.

(4) Choose the calculation method for spectral similarity.

(5) Click the "Run Mining-MS2 Level" button to perform matching mining.

### 4. High confidence compounds based on Sirius annotation

The "Tab-Annotation" automatically annotates high-confidence mined compounds and

further validates the matching results based on the predicted molecular formulae of the products.

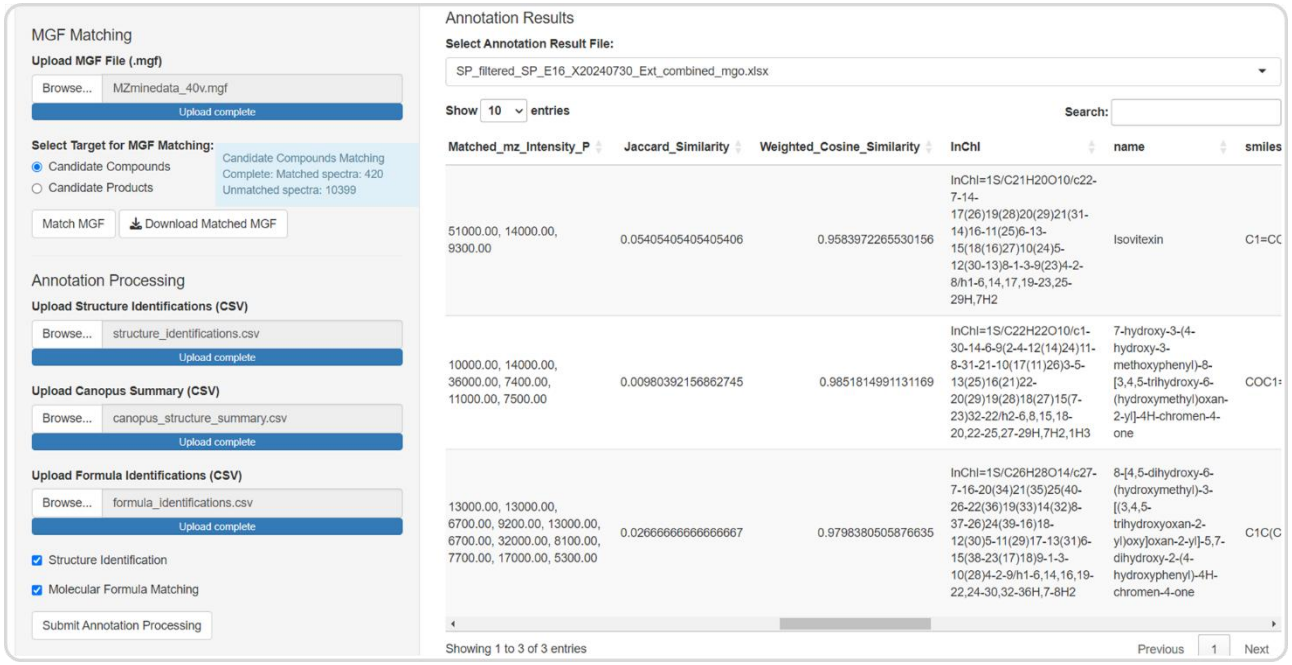


**Fig 5.** The Annotation function of the CTM-AM interface.

(1) Upload the .mgf file and select the substrate or product list that requires Sirius annotation.

(2) Click the "Match MGF" button to generate the corresponding file.

(3) Upload .csv files of substrates and products separately for matching and validation by obtaining Sirius' annotation list

(4) Click the "Submit Annotation Processing" button to view the annotation results.

# The Mining and Annotation workflow of the CTM-AM

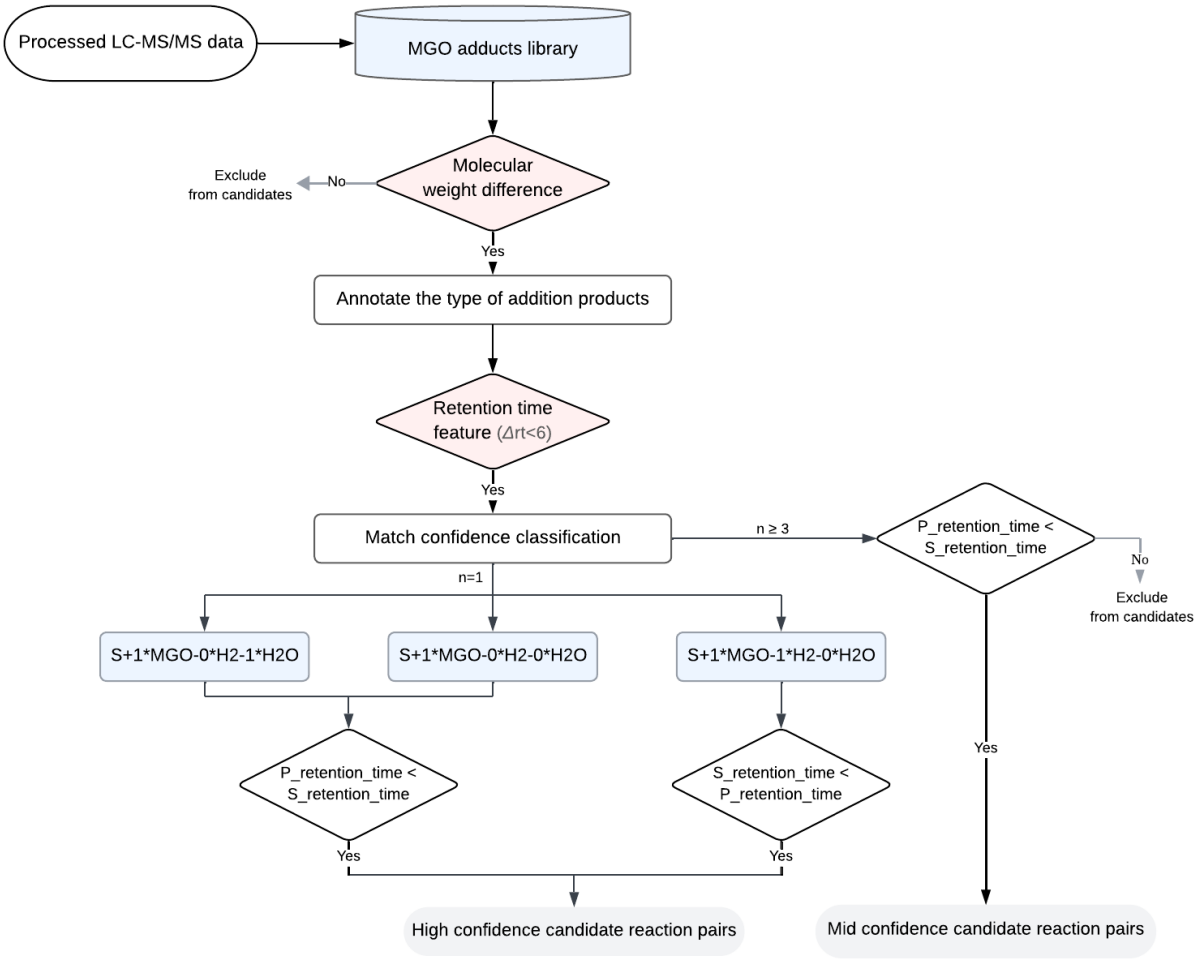


**Fig 6.** The CTM-AM mining workflow based on MS1 level.

# The Mining and Annotation workflow of the CTM-AM

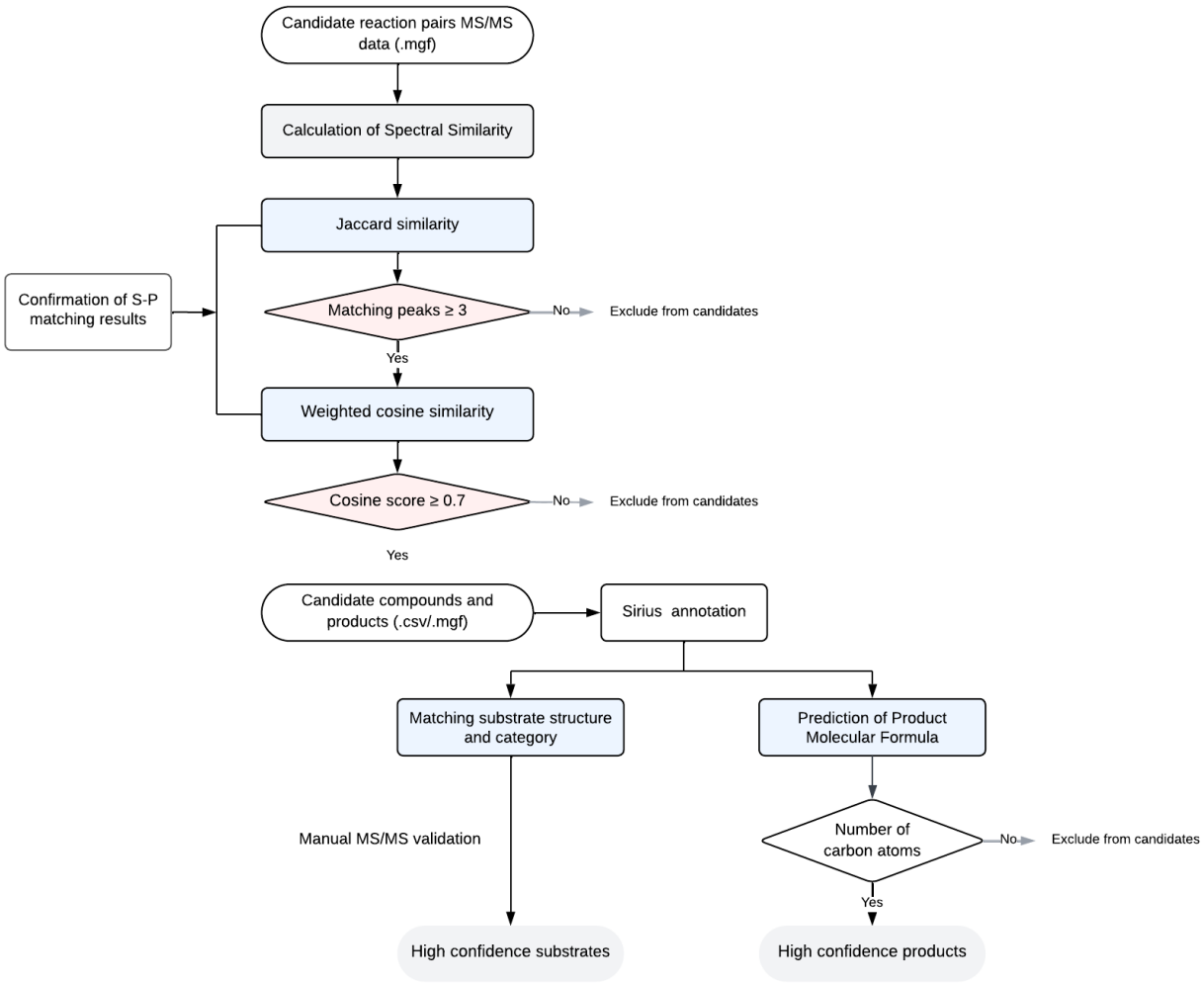


**Fig 7.** The CTM-AM mining workflow based on MS2 level and Sirius annotation.

# Result output list

## 1. LC-MS data of candidate active compounds and matching products

| S_value | S_retention_time | S_Peak.area1 | S_Peak.area2 | P_value | P_retention_time | P_Peak.area2 | n | value_type | match_label |
|---|---|---|---|---|---|---|---|---|---|
| 595.1650579 | 5.901824 | 10300000 | 0 | 667.1861251 | 5.191679 | 130796.65 | 1 | S+1*MGO-0*H2-0*H2O | High confidence |
| 595.1650579 | 5.901824 | 10300000 | 0 | 667.1866177 | 5.3211374 | 116961.41 | 1 | S+1*MGO-0*H2-0*H2O | High confidence |
| 611.1594097 | 6.022133 | 166022.88 | 0 | 683.1812551 | 6.143692 | 217598.48 | 1 | S+1*MGO-0*H2-0*H2O | Mid confidence |
| 611.1599373 | 5.435599 | 55600000 | 13200000 | 755.2017545 | 4.4442368 | 1340893.1 | 2 | S+2*MGO-0*H2-0*H2O | Mid confidence |
| 611.1599373 | 5.435599 | 55600000 | 13200000 | 755.2018358 | 4.6019 | 825835 | 2 | S+2*MGO-0*H2-0*H2O | Mid confidence |
| 611.1599373 | 5.435599 | 55600000 | 13200000 | 755.2021021 | 4.271996 | 638012.94 | 2 | S+2*MGO-0*H2-0*H2O | Mid confidence |
| 625.1755556 | 5.973773 | 12700000 | 1614444.4 | 697.1966832 | 5.416439 | 293388 | 1 | S+1*MGO-0*H2-0*H2O | High confidence |
| 625.1755556 | 5.973773 | 12700000 | 1614444.4 | 769.2181255 | 5.0921497 | 61245.45 | 2 | S+2*MGO-0*H2-0*H2O | Mid confidence |
| 645.1653011 | 4.33785 | 116950.8 | 0 | 753.1852486 | 5.1618557 | 97370.31 | 2 | S+2*MGO-0*H2-2*H2O | Mid confidence |
| 645.1653011 | 4.33785 | 116950.8 | 0 | 753.186065 | 5.022943 | 957535.25 | 2 | S+2*MGO-0*H2-2*H2O | Mid confidence |
| 645.1653011 | 4.33785 | 116950.8 | 0 | 753.186251 | 4.8009887 | 554811.1 | 2 | S+2*MGO-0*H2-2*H2O | Mid confidence |

(1) ***s value***: The molecular weight of the precursor ion of the reaction substrate.

(2) ***s retention time***: The retention time of precursor ions in the reaction substrate.

(3) ***s peak.area1***: The peak area of the reaction substrates after 24 hours without adding MGO.

(4) ***s peak.area2:*** The peak area of the reaction substrates after 24 hours of reaction with MGO.

(5) ***p value***: The molecular weight of the precursor ion of the products.

(6) ***s retention time***: The retention time of precursor ions in the products.

(7) ***s peak.area2***: The peak area of the product after 24 hours of reaction with MGO.

(8) ***n***: The number of MGO adducted to the substrate after the reaction.

(9) ***value type***: The types of MGO attached to substrates after reaction.

## 2. LC-MS data of candidate active compounds and matching products

| Matched_mz_S | Matched_mz_P | Matched_mz_Intensity_S | Matched_mz_Intensity_P | Jaccard_Similarity | Weighted_Cosine_Similarity |
|---|---|---|---|---|---|
| 57.0343, 61.0291, 69.034 | 57.0343, 61.0291, 69.0 | 1300000.00, 450000.00, 100( | 26000.00, 8900.00, 15000.0( | 0.025641026 | 0.997365638 |
| 57.0343, 61.0291, 69.034 | 57.0343, 61.0291, 69.0 | 1300000.00, 450000.00, 100( | 21000.00, 6100.00, 13000.0( | 0.026315789 | 0.996681572 |
| 53.0394, 55.0186, 57.034 | 53.0393, 55.0186, 57.0 | 48000.00, 60000.00, 290000.( | 5600.00, 6200.00, 48000.00, | 0.014925373 | 0.757247628 |
| 57.0343, 69.0342, 71.049 | 57.0343, 69.0341, 71.0 | 1600000.00, 910000.00, 120( | 56000.00, 16000.00, 340000 | 0.007751938 | 0.941683019 |
| 71.0499, 85.0290, 129.05 | 71.0499, 85.0290, 129 | 12000000.00, 18000000.00, 1 | 57000.00, 69000.00, 7900.0( | 0.022988506 | 0.949424656 |
| 55.0550, 57.0343, 61.029 | 55.0550, 57.0343, 61.0 | 160000.00, 1600000.00, 590( | 6500.00, 46000.00, 9900.00, | 0.014925373 | 0.950055182 |
| 53.0394, 55.0186, 57.034 | 53.0394, 55.0186, 57.0 | 190000.00, 200000.00, 1700( | 6700.00, 7000.00, 55000.00, | 0.014705882 | 0.994252097 |
| 71.0499, 85.0290, 92.007 | 71.0499, 85.0290, 92.0 | 13000000.00, 19000000.00, 7 | 43000.00, 61000.00, 8900.0( | 0.018018018 | 0.837692111 |
| 71.0499, 85.0290, 92.008 | 71.0498, 85.0291, 92.0 | 39000.00, 53000.00, 5100.00 | 9700.00, 15000.00, 5400.00 | 0.083333333 | 0.976067475 |
| 57.0342, 71.0499, 85.029 | 57.0342, 71.0499, 85.0 | 5600.00, 39000.00, 53000.00 | 120000.00, 710000.00, 9200 | 0.011627907 | 0.999732708 |
| 57.0342, 71.0499, 85.029 | 57.0343, 71.0499, 85.0 | 5600.00, 39000.00, 53000.00, | 55000.00, 380000.00, 57000 | 0.011111111 | 0.993942453 |

(1) ***matched mz s***: The m/z of the feature fragments matched in the MS2 spectrum of the substrate.

(2) ***matched mz p***: The m/z of the feature fragments matched in the MS2 spectrum of

the product.

(3) ***matched mz intensity s***: The peak intensity values of characteristic fragments matched in the MS2 spectrum of substrate.

(4) ***matched mz intensity p***: The peak intensity values of characteristic fragments matched in the MS2 spectrum of product.

(5) ***jaccard similarity***: The calculation value of Jaccard similarity between substrate and corresponding product.

(6) ***weighted cosine similarity***: The calculation value of weighted cosine similarity between substrate and corresponding product.

## 3. LC-MS data of candidate active compounds and matching products

| name | ConfidenceScoreExact | ConfidenceScoreApproximate | molecularFormula | adduct | NPC#class | P_molecularFormula | P_sirius_molecularFormula | P_label |
|---|---|---|---|---|---|---|---|---|
| 3-[[4,5-dihydroxy-6-(hydroxymethyl)-3-[( | 0.354 | 0.48 | C27H30O15 | [M+H]+ | Flavonols | C30H34O17 | C30H34O17 | high confidence |
| 3-[[4,5-dihydroxy-6-(hydroxymethyl)-3-[( | 0.354 | 0.48 | C27H30O15 | [M+H]+ | Flavonols | C30H34O17 | C30H34O17 | high confidence |
| Rutin | 0.649 | 0.718 | C27H30O16 | [M+H]+ | Flavonols | C30H34O18 | C30H34O18 | high confidence |
| Rutin | 0.776 | 0.848 | C27H30O16 | [M+H]+ | Flavonols | C33H38O20 | C33H38O20 | high confidence |
| Rutin | 0.776 | 0.848 | C27H30O16 | [M+H]+ | Flavonols | C33H38O20 | C33H38O20 | high confidence |
| Rutin | 0.776 | 0.848 | C27H30O16 | [M+H]+ | Flavonols | C33H38O20 | C33H38O20 | high confidence |
| Keioside | 0.63 | 0.745 | C28H32O16 | [M+H]+ | Flavonols | C31H36O18 | C31H36O18 | high confidence |
| Keioside | 0.63 | 0.745 | C28H32O16 | [M+H]+ | Flavonols | C34H40O20 | C34H40O20 | high confidence |
| 6-Hydroxy-Rutin | 0.311 | 0.582 | C27H30O17 | [M+H2O+H]+ | Flavonols | C33H34O19 | C33H36O20 | mid confidence |
| 6-Hydroxy-Rutin | 0.311 | 0.582 | C27H30O17 | [M+H2O+H]+ | Flavonols | C33H34O19 | C33H36O20 | mid confidence |
| 6-Hydroxy-Rutin | 0.311 | 0.582 | C27H30O17 | [M+H2O+H]+ | Flavonols | C33H34O19 | C33H36O20 | mid confidence |

(1) ***name***: The Sirus predicts candidate names based on database molecular fingerprints.

(2) ***confidence score exact***: The structure prediction score of Sirus in Exact mode.

(3) ***confidence score approximate***: The structure prediction score of Sirus in Approximate mode.

(4) ***molecular formula***: The molecular formula of the compound predicted by Sirus.

(5) ***adduct***: The adduct of the compound predicted by Sirus.

(6) ***class***: The structural categories of compounds predicted by CANOPUS.

(7) ***p molecular formula***: The molecular formula of the product obtained by matching the carbonyl product library.

(8) ***p sirius molecular formula***: The molecular formula of the product predicted by Sirus.

(9) ***p label***: Distinguish confidence levels by comparing predicted and matched molecular formulas.