**FragAssembler**

FragAssembler is an open-source cheminformatics program for elucidating the possible chemical structure of biotransformation products and reducing the number of false-positive structure candidates during the identification of xenobiotic biotransformation products.

1. **Software environments**
2. Windows OS: Windows 7 or later
3. RAM: 8.0 GB or more
4. Programming language: R (R 4.1.0 or later)
5. **Required programs**
6. FragAssembler programs (including biotransformation\_assigner and Fragment\_signature\_assembler)
7. MS-FINDER (Download link: <http://prime.psc.riken.jp/Metabolomics_Software/MS-FINDER/index.html>)
8. “ChemmineR” package (install command is involved in Fragment\_signature\_assembler program)
9. “lsa” package (install command is involved in Fragment\_signature\_assembler program)
10. **Required files**
11. Biotransformation database table (.csv)
12. Biotransformation product candidates list table (.csv)
13. Assigned biotransformation reactions data table (.csv)
14. MS/MS data of parent compound and biotransformation products (MAT format, the detail of MAT format can see the link: https://mtbinfo-team.github.io/mtbinfo.github.io/MS-FINDER/tutorial)
15. **Recommended data type and acquisition mode**

The FragAssembler is the tool to elucidate the possible chemical structure of biotransformation products and reduce the number of false-positive structure candidates during the identification. The calculation of FragAssembler is based on the MS/MS data of parent compound and biotransformation products. To acquire these MS/MS data, the in vitro or in vivo metabolism experiment and the untargeted metabolomics approaches could be used to obtain the sample with parent compound and biotransformation products and find the signals of these compounds. The detailed analysis approach can refer in our publication. For the MS/MS data acquisition, FragAssembler needs high-quality and high-resolution MS/MS data. The Q-tof or Orbitrap mass spectrometer with data-dependent acquisition (DDA) mode or targeted MS/MS acquisition mode is recommended to acquire the MS/MS data of interested compounds and biotransformation products.

1. **FragAssembler processing procedure**

The processing procedure, input data/parameters, and output result of FragAssembler are illustrated in Figure 1. The processing procedure involves 3 steps, biotransformation assignation (using biotransformation assigner), Structural elucidation of fragments of the parent compound (using MS-FINDER), and fragment signature assembling (using fragment signature assembler).

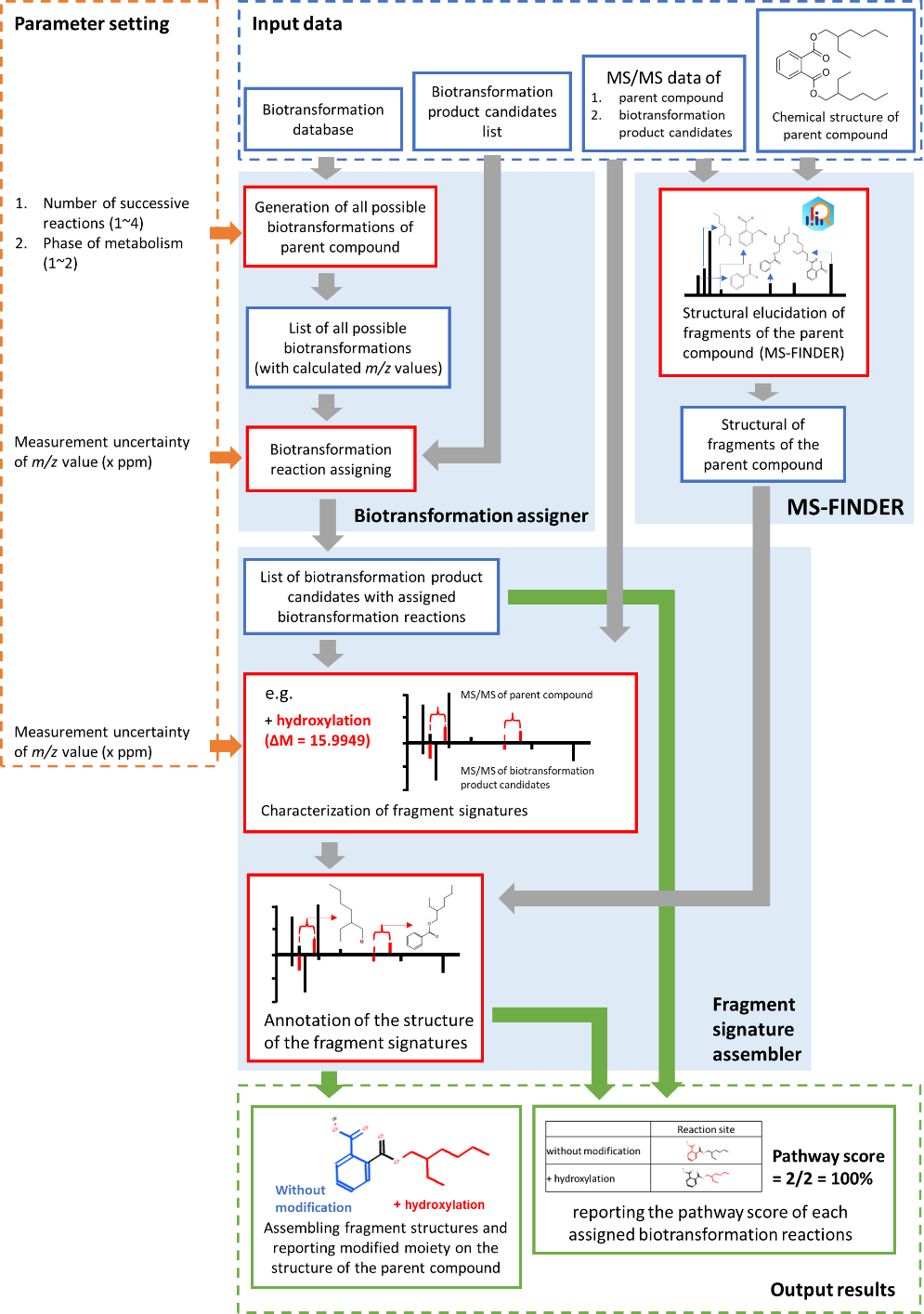


Figure 1. Schematic illustration of FragAssembler processing procedure

**5-1.** **Biotransformation assigner**

The biotransformation assigner has two steps: (1) generating all possible biotransformation reactions of the targeted parent compounds and (2) matching the possible biotransformation reactions to the biotransformation product candidates. Before the calculation, the data of biotransformation reactions (Biotransformation database table) and biotransformation product candidates (Biotransformation product candidates list table) should be input in the program using the commands:

meta\_pathy\_list <-as.matrix(read.csv(file.choose(), header=T, sep=",")) #input biotransformation database

meta\_cand\_list <-as.matrix(read.csv(file.choose(), header=T, sep=",")) #input biotransformation product candidates list

The format of these two data can refer to the example data. For the first step of the biotransformation assigner, the number of successive reactions and phase of metabolism need to be set using the commands:

roop\_num <- 4

phase <- 1

After generating all possible biotransformation reactions of the targeted parent compounds, these calculated reactions will match the biotransformation product candidates list in the biotransformation product candidates list table with the set mass error:

ppm\_data <- 10

The assigning of possible biotransformation reactions to biotransformation product candidates can be based on mass error or molecular formula. The calculation result base on mass error is exported as "Assigning\_biotransformation\_reaction\_mass.csv" and the resulting base on molecular formula is exported as "Assigning\_biotransformation\_reaction\_formula.csv" in the "Documents" folder.

**5-2.** **Input fragments information of parent compound base on MS-FINDER**

For the annotation of possible moiety with biotransformation on the parent compound, the structure of each fragment of the parent compound needs to be inputted into FragAssembler. The in silico fragment annotation tool MS-FINDER is recommended to apply in the calculation of FragAssembler. Other in silico fragmentation tools, such as CFM-ID[1], MetFrag[2], and Mass Frontier™ (Thermo Fisher Scientific), can also be applied in this program. For the first step of fragment calculation, input the MS/MS data of the parent drug (.mat format) in the MS-FINDER software and calculate the in silico fragment annotation result of the parent drug (Figure 2).

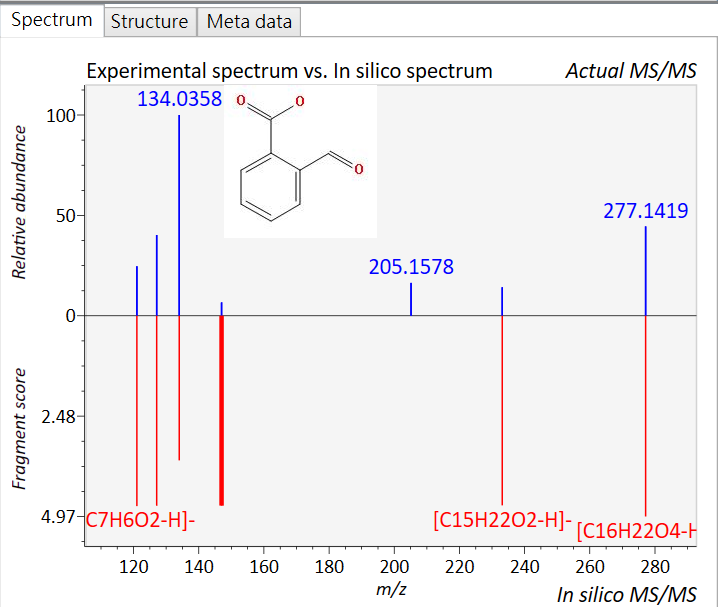


Figure 2. *in silico* fragment annotation result of parent drug based on MS-FINDER. Take mono(2-ethylhexyl)phthalate (MEHP) as the example.

The detail of using MS-FINDER software can be found at <https://mtbinfo-team.github.io/mtbinfo.github.io/MS-FINDER/tutorial>. For the second step, open the “Fragment\_signature\_assembler” program of the FragAssembler and install the ChemmineR package using the following commands:

if (!requireNamespace("BiocManager", quietly=TRUE))

install.packages("BiocManager")

BiocManager::install("ChemmineR")

After that, input the structure of the parent compound using the following commands:

pd\_structure\_smile <- read.SDFset("/Structure.sdf")

plot(pd\_structure\_smile, atomnum = TRUE ,no\_print\_atoms = "",atomcex=0.7, print=FALSE)

The "/Structure.sdf" is the .sdf file of the structure of the parent compound, which can be acquired in the PubChem online database or generated on the OpenBabel website (<http://www.cheminfo.org/Chemistry/Cheminformatics/FormatConverter/index.html>). The “plot” command can draw the structure and label each atom of the parent compound (Figure 3). The parameter "atomcex=0.7" can use to change the size of the label number.

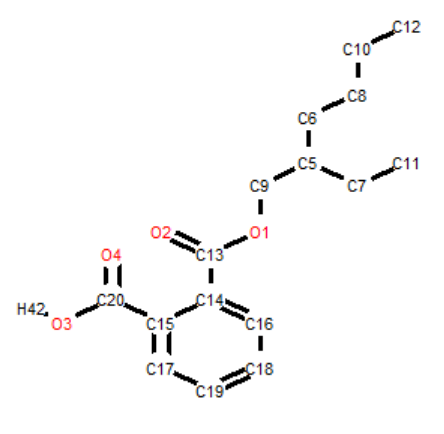


Figure 3. the illustrated structure and labeled atoms of parent compound drawn by ChemmineR package. Using the MEHP as the example.

For the third step, the structure information of parent compound fragments should be input into the FragAssembler. In line 88 of“Fragment\_signature\_assembler” program, fragments are inputted based on these commands labeled with #manual input. These commands should be modified when the FragAssembler is calculated for different parent compounds because the fragments of different parent compounds are also different. The detailed fragment input progress is illustrated in Figure 4.

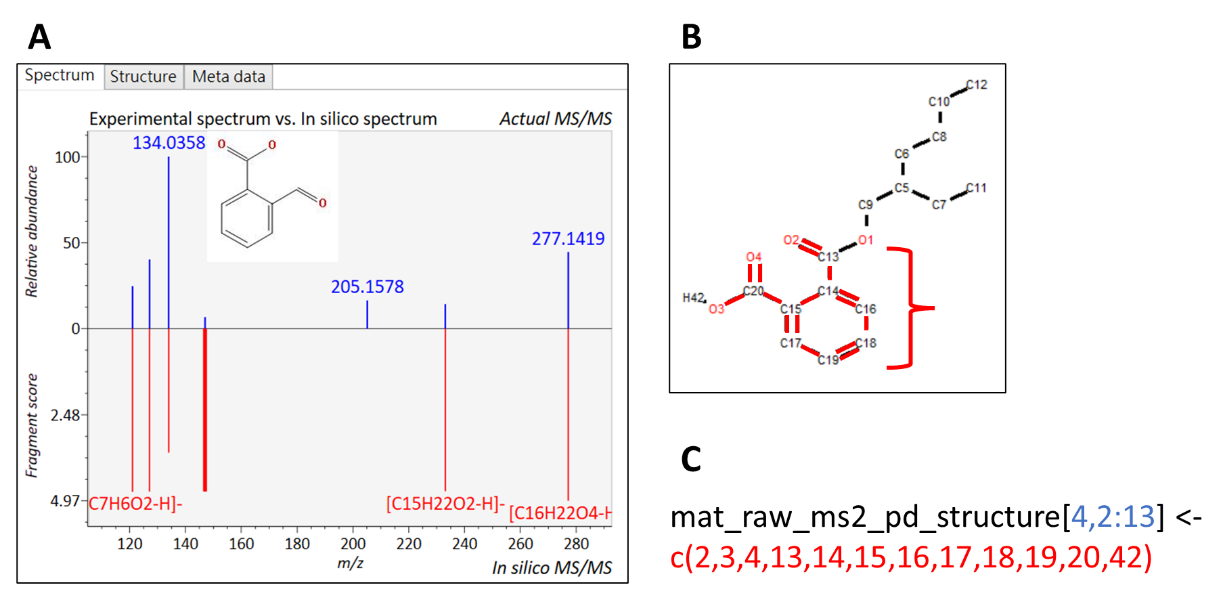


Figure 4. fragment input progress of FragAssembler. Take 4th fragment of MEHP as an example.

The blue number in Figure 4C, “4”, symbolizes the 4th fragment of the targeted parent compound. The “2:13” is the serial number of labeled atoms of fragments. In the example in Figure 4, the fragment has 12 atoms, so the series number of labeled atoms is “2:13”. The red code in Figure 4C is the site of fragment atoms. The fragment structure annotation result (Figure 4A) can be mapped on the structure generated by ChemmineR (Figure 4B) and labeled in the code (Figure 4C). The parent compound with multiple fragments could lead to a more tedious input process. The setting of cutoff value in mass error or fragment intensity is recommended to reduce the cumbersome of the fragment input process.

**5-3.** **Fragment signature assembler and data output**

The “Fragment\_signature\_assembler” program would calculate the possible moieties with biotransformation corresponding to each biotransformation product candidate. Three types of data should be inputted into the program, the fragment of the targeted parent compound, MS/MS data of the parent compound and biotransformation product candidates, and the possible biotransformation reactions list of biotransformation product candidates. The inputting of the first data is described in sections 5-2. The second and third data are included in the file “Assigned biotransformation reactions data table.” The table includes the file path of each MS/MS data and calculated possible biotransformation reactions (each biotransformation product candidate has 10 reaction candidates). The format of the “Assigned biotransformation reactions data table” can see in the example file.

After input all the required data, the fragment signature assembler can be automatically processing and generate the results. The calculation result involves the modified moiety and the score of each biotransformation product candidate. The modified moiety data is in “metabolite id result position f1~f10” table. The “f1~f10” represents the 10 reaction candidates in “Assigned biotransformation reactions data table”. The export procedure of modified moiety data is illustrated in Figure 5.

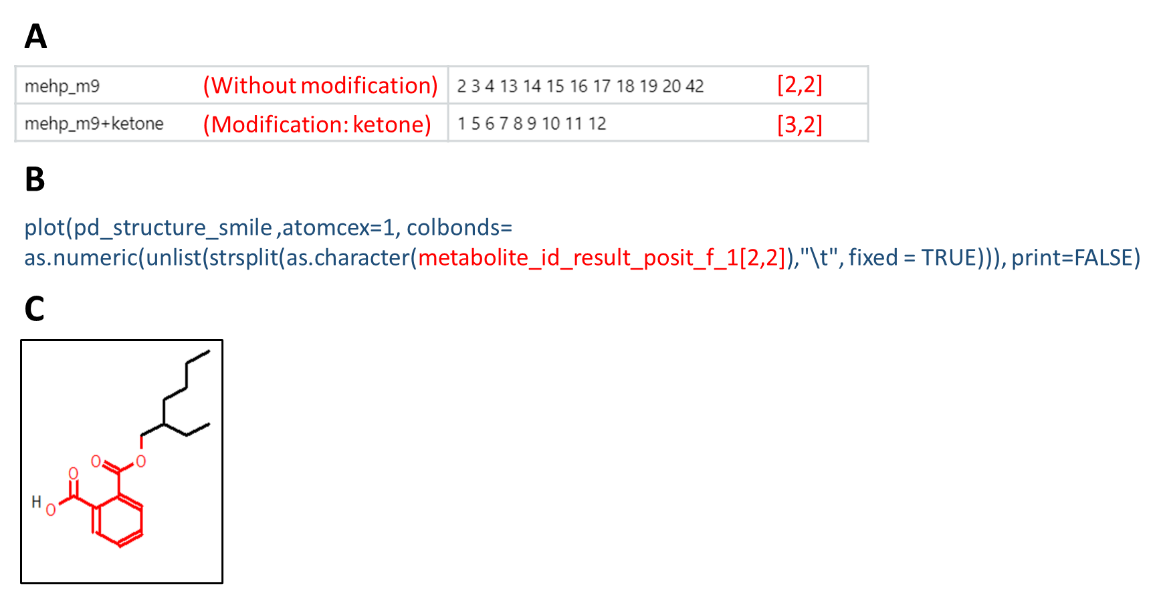


Figure 5. export procedure of modified moiety data.

One “metabolite id result position” table involves all possible modifications of all biotransformation product candidates based on the one reaction candidate “Assigned biotransformation reactions data table.” The modified moiety calculation result of one biotransformation product candidate in the “metabolite id result position” is like Figure 5A. In this example, reaction candidate “ketone” has two possible modifications: “without change” and “Modification: ketone,” which are labeled as “m19” and “m19 + ketone”. The number series in the next column is the calculation result representing the modification on the left. The command in Figure 5B can be used to visualize the result. For example, we want to look at the “metabolite\_id\_result\_posit\_f\_1[2,2]” position. We can enter the command in Figure 5B, and the R studio program will show the result in the “Plots” interface.

The score results are in the “metabolite\_id\_result\_table\_score\_p” table. Each row represents the biotransformation product candidate, and each column represents 10 reaction candidates. The detail of the score can be found in the publication.

Getting Help

If you have any questions about this notebook, please email tp6m0654@gmail.com

References

[1] Allen F, Pon A, Wilson M, Greiner R, Wishart D. CFM-ID: a web server for annotation, spectrum prediction and metabolite identification from tandem mass spectra. Nucleic Acids Res 2014;42:W94-W9.

[2] Ruttkies C, Schymanski EL, Wolf S, Hollender J, Neumann S. MetFrag relaunched: incorporating strategies beyond in silico fragmentation. Journal of cheminformatics 2016;8:1-16.