User manual for deconvolution and ULSA for data independent HRMS.

The data, library, and feature list structure.

Before starting with the computation, it is extremely important to configure your data (i.e. chromatogram) and your local reference library in the following format.

Chromatogram:

The **mass calibrated chromatogram** must be stored as a "mat" file being a matlab structure data file. Please make sure the field names in the files are kept as the example file "Chrom". Below you can see the structure of the chromatogram acceptable to the function "DeconPlusULSA". The main field in this structure file must be called "data_c" for mass calibrated data.

data_c =

struct with fields:

Mz_values: [1859×52331 double]

Mz_values_2: [1858×34392 double]

Mz_intensity: [1859×52331 double]

Mz_intensity_2: [1858×34392 double]

In this file "Mz_values" represents the measured m/z values at the lowest collision energy, "Mz_values_2" represents the measured m/z values at high collision energy, "Mz_intensity" represents the intensity for each measured m/z value at the low collision energy, and finally, "Mz_intensity_2" represents the intensity for each measured m/z value at the high collision energy. Each row in a field for this data structure is a scan while each column is a mass channel. In other words, in the example file the mass spectrometer recorded 1,859 scans and 52,331 mass channels for low energy channel and 34,392 mass channels for high energy channel. **Please make sure that the two energy channels are well aligned before starting your computations.**

Reference Library:

The current implementation of this workflow uses a local version of MassBank (present in the main folder). If you want to use another spectral library, you need to configure your library in the following format for the function (i.e. "DeconPlusULSA") to read it. In this case also the whole library is stored as structure in ".mat" format (see below). In this data file each column represents a library entry. The main field in this file is called "**data**". Please make sure that the field names are kept as is shown below. Otherwise the main function (i.e. "DeconPlusULSA") would not be able to read the library file.

data =

mz_values: [2210×46017 double]

mz_Int: [2210×46017 double]

mz_rel_Int: [2210×46017 double]

Exact_mass: [2×46017 double]

Name: {1×46017 cell}

Ion_mode: {1×46017 cell}

Ion_source: {1×46017 cell}

Smiles: {1×46017 cell}

InChi: {1×46017 cell}

CAS: {1×46017 cell}

Splash: {1×46017 cell}

Access: {1×46017 cell}

Date: {1×46017 cell}

MS_Type: {1×46017 cell}

CE: {1×46017 cell}

Chenical_Formula: {1×46017 cell}

Feature list:

The feature list is an excel file with three columns: “ID”, which is a number given to each feature and the result file will reference to that number; “Retention_time”, which is the scan number of the feature to be identified; and “MZ”, which is the measured mass of that feature. An example of this file is provided in the main folder called “Targets_1”. Please make sure that the headers remain the same for your future calculations.

Getting started:

Once you have the raw data, reference library, and the feature list configured to be read by the main function, you are ready to start the calculations. To start your calculations, you need to take the following steps:

1. Change the current working directory in matlab to the main folder.
2. Open the file called “DeconPlusULSA”.
3. Set all the parameters in that file. The description of each parameter is provided in that file. Please read the descriptions properly and make sure you understand it.
4. Hit run.

Results:

The main function creates three nested folders:

a. Named based on the parameter “Project_name”
   a.1. Named based on the parameter “File_name_IDs”
   a.2. Named based on the parameter “File_name_Spec”

In the first folder you will find a txt file for each feature in your feature list. This file contains all the candidates that are relevant to that specific feature. The second folder contains txt files with the extracted pseudo MS2 spectra for each feature. If there is no file recorded for a feature, this implies that the algorithm was not successful in extracting the spectra or identifying that feature.

One of the most important parameters to optimize when using this algorithm is the parameter “Weigth_fun”. This parameter defines the importance of each of seven parts of the final score calculation. Please see:

Combining a Deconvolution and a Universal Library Search Algorithm for the Nontarget Analysis of Data-Independent Acquisition Mode Liquid Chromatography-High-Resolution Mass Spectrometry Results", Saer Samanipour, Malcolm J. Reid, Kine Bæk, and Kevin V. Thomas, 2018, ES&T, 10.1021/acs.est.8b00259

for detailed description of these seven parts. The “Weigth_fun” set to one for all seven parts produces a final score of 7 for a perfect match and a score of 0 for two orthogonal spectra. When trying to set this parameter (i.e. “Weigth_fun”), it is recommended that you optimize it using a well-known feature. This will enable you the set the threshold on the acceptable “Score” values. When interpreting the final results, it is really important to make sure that at least three fragments are matched.

Contact info:

Please get in touch if you need help or would like to collaborate.

Dr. Saer Samanipour

Norwegian Institute for Water Research (NIVA)

NIVA, Gaustadalleen 21, No-0349 Oslo

Mob: +47 98222087

Email: saer.samanipour@niva.no

www.niva.no