

Introduction

Molecular networking (GNPS) and mass spectrometry (MS) analysis in general are sensitive to the choice of software parameters. While the [originally published](#) analysis OF MSV000084900 is available (<https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=d5adba7f67cc402396e9ba7cd85ce52b>) it was performed on 2020-02-10 which means its molecular network is too large to open (because the graphml file contains all the singleton clusters) and the library search is out of date. Therefore, a new run using the new GNPS2 and updated libraries was performed on 2025-03-23 (<https://gnps2.org/status?task=18d82f1067e643538adf9b73147122c3>) and downloaded to ([data/gnps2_results/18d82f1067e643538adf9b73147122c3](#)); lastly, a custom search was performed using custom scripts and [matchms](#). The scripts and notebooks in the analysis directory synthesizes all of these results. Parameters used for NPS2 analysis can be found in [data/gnps2_results/18d82f1067e643538adf9b73147122c3/submission_parameters](#).

GNPS molecular networking / what is a “cluster”

GNPS (Global Natural Products Social Molecular Networking) molecular networking is a computational approach used in mass spectrometry (MS) to visualize and analyze the similarity of chemical compounds based on their fragmentation patterns.

In LC-MS/MS, samples first pass through a column that physically separates molecules based on their physicochemical properties. In the case of dataset MSV000084900, a C18 column is used, which primarily separates compounds based on hydrophobicity. Depending on their strength of association with the column, molecules elute into the mass spectrometer (mass spec) at different times. The mass spec measures all masses within a specified range x times per second in MS1 scans (each measurement is called a scan). For each MS1 scan, a list of the n -most intense ion peaks (precursor ions) is generated. Over the next n scans, one of these precursor ions is selected, fragmented, and its fragment masses are measured in MS2 scans.

Thus, every molecule analyzed (in theory) has:

- A precursor ion (seen in an MS1 scan), which provides the mass of the intact molecule.
- A fragmentation fingerprint (seen in an MS2 scan).

GNPS2 “classic” molecular networking uses MS-Cluster to cluster paired MS1 precursor and MS2 fragment spectra based on their precursor mass and the cosine similarity of MS2 fragmentation spectra. Each cluster represents, in theory, a single molecular structure (e.g., caffeine).

However, the situation is more complex. A single molecule (e.g., caffeine) can appear in multiple clusters because molecules can exist in different ionization states in the mass spectrometer. Additionally, molecules can form adducts (e.g., molecule + solvent molecule), leading to variations in measured masses.

If you are familiar with GNPS network visualizations, these clusters correspond to the circles (nodes) in the network. The consensus MS2 spectrum for each cluster is then compared to all others, and clusters are linked based on their MS2 similarity.

Reproducibility

The notebooks and results were generated with Docker images that take advantage of [Pixi](#) environments. All computation was performed on an AMD Ryzen™ 9 7950X with 96.0 GiB DDR5 RAM, running Ubuntu 24.04.2 LTS and Docker version 28.0.1, build 068a01e. Software used is specified in `./docker/Dockerfile` and `pixi.toml`, with locked versions in `pixi.lock`

Get large data

In the top analysis directory, run:

```
# "Date Last Libraries Fully Exported - 2025-03-23 18:06:56.154000"
wget https://external.gnps2.org/gnpslibrary/ALL_GNPS.mgf -O data/ALL_GNPS.mgf
```

```
wget -r -np -nH -P data2 ftp://massive-ftp.ucsd.edu/v02/MSV000084900/
```

Create Docker image

In the top analysis directory, run:

```
docker build -t metabolite_search ./docker
```

Render notebooks

In the top analysis directory, run:

```
docker run -it --user $(id -u):$(id -g) \
  -v "$PWD":/metabolite_search \
  --workdir /metabolite_search metabolite_search \
  "quarto render 1_gnps2_molecular_networking.qmd"

docker run -it --user $(id -u):$(id -g) \
  -v "$PWD":/metabolite_search \
  --workdir /metabolite_search metabolite_search \
```

```
"quarto render 2-1_custom_search.qmd"

docker run -it --user $(id -u):$(id -g) \
  -v "$PWD":/metabolite_search \
  --workdir /metabolite_search metabolite_search \
  "quarto render 2-2_custom_search.qmd"

docker run -it --user $(id -u):$(id -g) \
  -v "$PWD":/metabolite_search \
  --workdir /metabolite_search metabolite_search \
  "quarto render 3_conclusion.qmd"
```