SUPPLEMENTARY MATERIAL

A general procedure for rounding *m/z* values in low resolution mass spectra

Mikhail Khrisanfov, Andrey Samokhin

Chemistry Department, Lomonosov Moscow State University, Leninskie Gory 1-3, 119991 Moscow, Russia.

How to convert NetCDF to CSV

AMDIS (NIST):

- Open a NetCDF file (‘File’ > ‘Open’).
- Select ‘Export TIC (text)’ from the ‘File’ menu.

ChemStation (Agilent):

- Import a NetCDF file (‘File’ > ‘Import AIA Raw Data Files’).
- Select ‘Export 3D Data’ from the ‘Tools’ menu.

ChromaTOF (LECO):

- Import a NetCDF file (‘File’ > ‘Import’).
- Right-click on the chromatogram, select ‘Export’ from the context menu.

OpenChrom (Lablicate):

- Open a NetCDF file (‘File’ > ’Open Chromatogram (MSD)’).
- Save CSV Chromatogram (‘File’ > ’Save as…’ with file type ‘CSV Chromatogram’).

Description of the Python scripts

- CSV table is loaded into Pandas DataFrame and preprocessed (any unneeded information is dropped) in function “reverse_algorithm” in file reverse.py.
- NetCDF is read using Xarray and split into separate NumPy arrays in function “read_netCDF” from “netCDF2csv.py”: “mass_output” is a two-dimensional array that contains arrays of floating-point *m/z* values for each of the scans; “abundance_output” contains abundances for the respective *m/z* values.
- SciPy function “scipy.optimize.minimize_scalar” is used to find algorithm for rounding most closely matching the results of one used in software in question. It changes bin border in algorithm for rounding until minimum Euclidean distance between DataFrame rounded by candidate algorithm for rounding and CSV table is found:
    - Floating-point *m/z* values in “mass_output” are rounded to integers using numpy.ceil() function: MZ = [mz – delta], where MZ and mz are *m/z* values; MZ is an integer number, mz is a floating-point number, delta is a floating-point number which is changed by “minimize_scalar”.
    - “Abundance output” and rounded “mass_output” are combined to form a new Pandas DataFrame.
    - Euclidean distance between candidate DataFrame and CSV table is calculated using “numpy.linalg.norm” fuction.

Description of the CSV files

- sample10.cdf –NetCDF file
- SAMPLE10.TIC – AMDIS export of sample10.cdf
- sample10_ChemStation.csv – ChemStation export of sample10.cdf
- sample10_ChromaTOF.csv – ChromaTOF export of sample10.cdf
- sample10_OpenChrom.csv – OpenChrom export of sample10.cdf
- sample10_MSSearch_def.csv – MS Search export (with default settings)
- sample10_MSSearch_M149.csv – MS Search export (with "Add this term to all m/z" = -0.149)

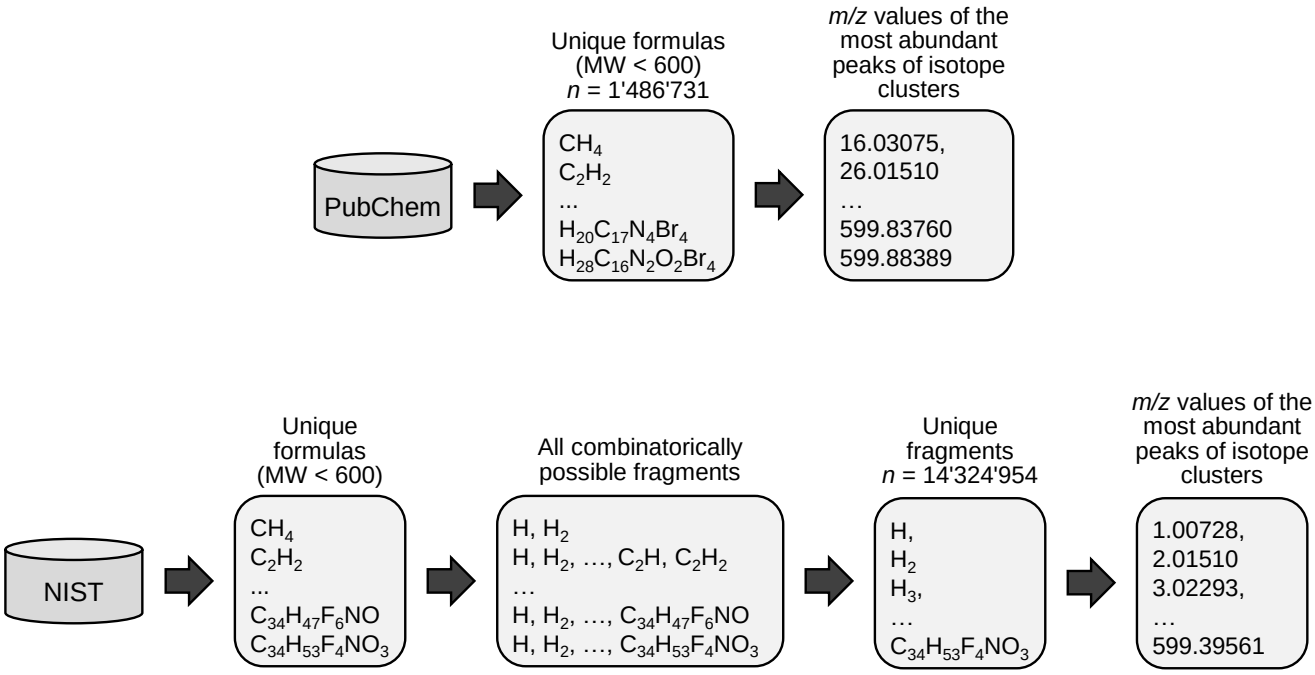

Fig. S1. Two model datasets based on PubChem and NIST databases.