mzCloud, a first quantum chemically annotated mass spectral database

Michal Raab1, Chongle Pan2, Milos Suchy3, Martin Mandak1, Juraj Lutisan1, Jakub Mezey1, Robert Mistrík4
1HighChem, Bratislava, Slovakia; 2Oak Ridge National Laboratory, Oak Ridge, TN

Introduction

Traditionally, a species' linear search technique using reference mass spectra is one of the most straightforward methods of compound identification in metabolomics. However, existing spectral libraries cover only a small portion of the compounds found in biological samples and even if they contain those compounds, the inherent spectra reproducibility problems in high-resolution LC/MS experiments hinder the identification further. To address the problem of reliable compound identification that plagues metabolomics applications, we have developed a freely accessible mass spectral database that contains experimental mass spectral data with high-quality fragment annotations derived from quantum-chemical (QC) calculations. This provides the essential means for the identification of unknowns when traditional library search methods are unsuccessful.

Methods

The annotation of the precursor ions and observed fragment peaks with corresponding molecular structures is carried out using heuristic and systematic bond cleavage methods, as implemented in the Mave, Frontier, and MeDMS algorithms. Predicted precursor ion abundances are then further processed in a progressively iterative QC computation to provide a reliable and chemically sound basis for the reconstruction of unknown compounds.

mzCloud in numbers

Unique compounds
5,000
Processed MS spectra
1,300,000
Peak annotations by formulae
6,600,000
Peak annotations by structures
500,000
QC calculated unique optimized structures
430,000

Results

The rationale behind the presented large-scale annotation of thousands of precursor ions and several million fragment peaks present in the mZCloud database (mzCloud.org) is the unique strategy of identification by connecting spectra from high-random MS stages. Precursor ion fingerprinting (PIF), PIF relies on well-defined and chemically plausible structures of fragment ions, which are either used to represent the parent compound or, at the very least, point towards its structural characteristics. This is of tremendous value in common cases when traditional library mining yields no results, and the analysis is left with absolutely no structurally known information.

To date, the mZCloud database features over 1,500,000 processed spectral records covering a wide range of colloid energies up to MSP 0.5 kV, human indigestible metabolites, plant secondary metabolites, food additives, pharmaceuticals, and other compounds relevant for metabolomics, and new records are accumulating rapidly. More than 100,000 individual processing pipelines for precursor ion prediction have been generated in order to find the best descriptor for each compound, resulting in the identification of 430,000 unique 3D structures with calculated thermocromical properties for more than 1.3 million MS spectra.

Conclusions

The mZCloud database provides both high-quality spectral collections and carefully selected peak annotations employing advanced QC methods, supplementing experimental data with independent theoretical predictions enabling a fundamentally different and thus unbiased, approach. As such, it is a time-saving and cost-effective tool extending its applications well beyond traditional library searches. Such tools are urgently needed to keep pace with time-critical analytical instrumentation that generates far more data than we are currently able to interpret.

Acknowledgements

We would like to thank Dr. Jonathan C. Sylve and Tim J. Strum from Thermo Fisher Scientific, for constructive discussions on the theoretical and experimental aspects of ion dissociation.