Fragmentation trees from mass spectra: a step towards metabolite identification

Florian Rasche, Aleš Svatoš, Ravi Kumar Maddula, Christoph Böttcher, Sebastian Böcker
{florian.rasche,sebastian.boecker}@uni-jena.de

The structural elucidation of organic compounds in complex biofluids and tissues remains a significant challenge. For mass spectrometry, the manual interpretation of tandem mass spectra is cumbersome and requires expert knowledge, as the fragmentation mechanisms of small molecules are not completely understood. Thus, the automated identification of compounds is generally limited to searching in spectral libraries.

We have developed an automated method for interpreting the fragmentation spectra of the organic compound, by computing fragmentation trees that establish not only the molecular formula of the compound and all fragment ions, but also dependencies between fragment ions. This is an important step toward the automated identification of unknown compounds that are not in any database.

The identification of small compounds from complex biofluids is crucial for many areas of biology and medicine, such as metabolomics and biomarker discovery. Due to its high throughput and sensitivity, mass spectrometry is commonly used for this task. The manual interpretation of fragmentation mass spectra is cumbersome and requires expert knowledge. Automated library search can only recover compounds already known, and thousands of metabolites for each species remain unidentified. Despite notable efforts since the 1960's as part of the DENDRAL project, very little progress has been made for the automated analysis of tandem mass spectra from these unknown compounds.

We have developed a method to automatically interpret tandem mass spectra of small compounds using fragmentation trees. These trees establish the molecular formula of the compound and all fragment ions, as well as the fragmentation reactions that created these fragments. Although computing a fragmentation tree is NP-hard, we developed an efficient yet exact algorithm for its solution, based on parameterized algorithmics. An extensive evaluation of fragmentation trees computed from reference data, showed that these trees agree very well with expert annotation of the data.

Our technique permits a comprehensive analysis well beyond what has previously been possible, since information about fragmentation reactions simplifies the manual analysis of tandem mass spectra. In the future, we may classify an unknown compound by comparing its fragmentation tree to reference trees. Another future application is the reconstruction of metabolic networks, where we can use fragmentation tree similarity to propose a new metabolic reaction.