An automated Metabolite Identification Pipeline using Mass Spectral Trees

Theo Reijmers[1], Julio Peironcely[1,2], Miquel Rojas-Chertó[1], Piotr Kasper[1], Leon Coulier[2], Albert Tas[2], Rob Vreeken[1], Thomas Hankemeier[1]

t.reijmers@lacdr.leidenuniv.nl

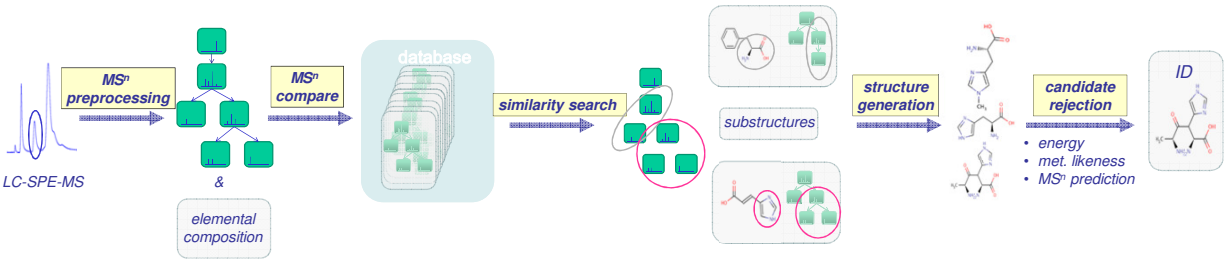
Netherlands Metabolomics Centre,
[1]Analytical Biosciences, Leiden University, [2]TNO Quality of Life, Zeist

Netherlands Metabolomics Centre

Introduction

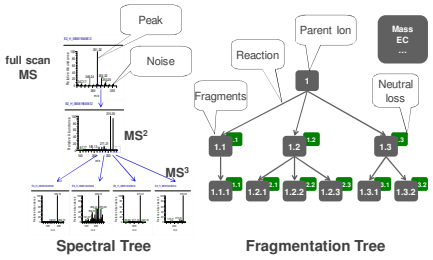
Identifying metabolites has been reported as one of the major bottlenecks in metabolomics. In part, this is due to the absence of good computational tools to automate metabolite identification. To address this issue, we have developed mathematical tools to process and compare multi-stage mass spectrometry data (MS^n), in order to extract as much as possible information from the fragmentation trees. In addition, candidate structures are generated computationally and filters are used to reject improbable chemical structures. This poster presents the integrated use of these tools in a pipeline fashion to identify metabolites in human urine.

Approach / Metabolite Identification Pipeline



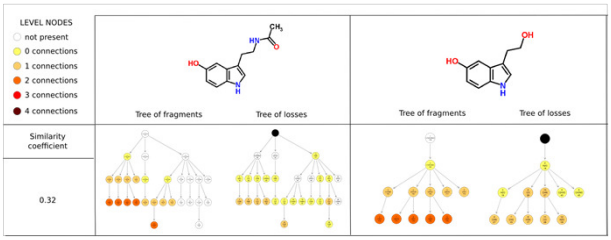
- MS^n spectral trees are processed into fragmentation trees using the MEF (1) tool. The nodes of the fragmentation and neutral loss tree are annotated with unique elemental compositions.
- Comparing the fragmentation tree with trees in a reference MS^n database (2) returns an identical tree (identity search) or multiple similar trees (similarity search). From the most similar trees maximum common substructures are extracted.
- The Open Structure Generator (OMG) (3) generates for the unknown metabolite all possible chemical structures for a given elemental composition and (multiple) substructures.
- This number of structures on the list of candidate structures is reduced after applying an internal energy and a metabolite likeness (4) filter. The fragmentation prediction tool MetFrag (5) further reduces the list by comparing the observed with predicted fragmentation spectra.

MS^n Preprocessing



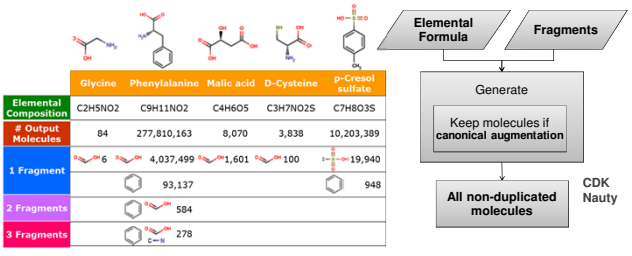
From a pooled urine sample, mass spectral trees are acquired for 30 compounds with unknown identity. After preprocessing with the MEF tool, 30 fragmentation and neutral loss trees are obtained with unique elemental compositions assigned to the nodes.

MS^n Comparison



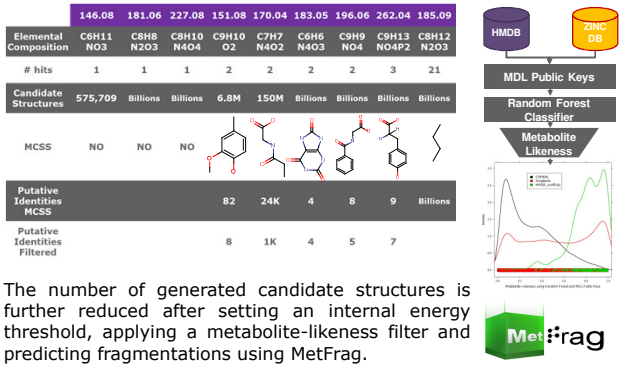
Similarity search in the NMC reference MS^n database of each unknown fragmentation tree yielded the following results: 9 trees are 100% similar to reference trees in database (identity search), 9 trees return similar trees from database (similarity search) and 12 trees do not show similar trees (similarity value < 10%).

Structure Generation



Multiple similar trees are found for 6 of the 9 trees. Maximum common substructure (MCSS) of these similar hits is used as additional input for structure generator. Remaining 3 trees only return 1 similar hit in the database so no MCSS can be calculated.

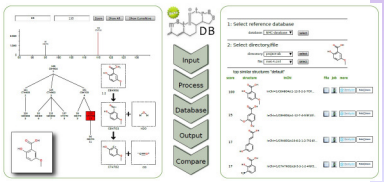
Candidate Rejection



The number of generated candidate structures is further reduced after setting an internal energy threshold, applying a metabolite-likeness filter and predicting fragmentations using MetFrag.

Conclusions

- **Presented pipeline, consisting of a number of separate tools, facilitates experts for (de-novo) identification of metabolites.**
- **Establishing a comprehensive MS^n database (for extracting substructures) is crucial.**
- **Visit the MetiTree (6) web-application, www.metitree.nl, for part of this pipeline.**



1. Rojas-Chertó, M. et al. *Bioinformatics* 27, **2011**, 2376-2383.
2. Rojas-Chertó, M. et al. *Anal. Chem.* 84, **2012**, 5524-5534.
3. Peironcely, J.E. et al. *J. Cheminformatics*, 4, **2012**, 21.
4. Peironcely, J.E. et al. *PLoS ONE*, 6, **2011**, e28966.
5. Wolf, S. et al. *BMC Bioinformatics*, 11, **2010**, 148.
6. Rojas-Chertó, M. et al. *Bioinformatics*, doi: 10.1093/bioinformatics/bts486.