

Enhancing Compound Identification for Untargeted Ion Mobility-MS Workflows

Aivett Bilbao¹, Joon Y. Lee¹, Hania Khouri², Thomas O. Metz¹, Richard D. Smith¹, Erin S. Baker¹, John Fjeldsted² and Samuel H. Payne¹

1) Biological Sciences Division, Pacific Northwest National Laboratory, Richland, WA, USA.

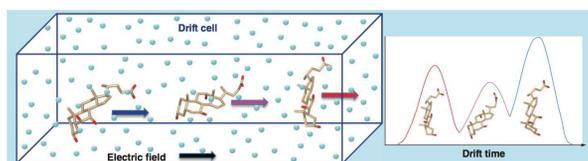
2) Agilent Technologies, Santa Clara, CA, USA

Overview

- Metabolite identification using collision cross section (CCS).
- Informatics pipeline for ion mobility-mass spectrometry-based study of the alpha-proteobacterium *Rhodospseudomonas palustris*.
- Similarity metrics for scoring metabolite features by combining CCS with other dimensions.

Introduction

- Collision cross section (CCS) is a powerful physicochemical property measurable by ion mobility mass spectrometry (IM-MS).
- CCS provides an additional and orthogonal dimension differentiating conformers otherwise undistinguishable by conventional separations (i.e., liquid or gas chromatography).

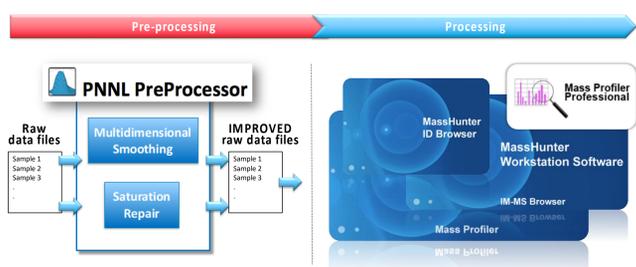


In ion mobility (IM), ions are separated based on their size and shape. In the example shown, the bile acids β -muricholic, urocholic and α -muricholic separate in order of fastest to slowest based on relative gas-phase ion shape [1].

- CCS is currently utilized as a simple tolerance filter or in combination with other dimensions (e.g., accurate mass, retention time and MS/MS) by means of a weighted sum of all scores where the user has to manually specify the weight/relevance of each one.

Methods

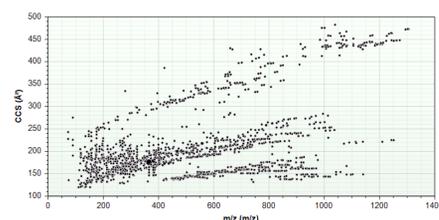
- Samples in three different growth conditions: aerobic, anaerobic nitrogen fixing and anaerobic photosynthetic.
- Direct infusion and ion mobility-mass spectrometry analyses on an Agilent IM-Q-TOF MS system with positive ionization.



- PNNL PreProcessor generated new MassHunter (.d) data files that can be processed with all other existing software.
- Free download: <https://omics.pnl.gov/software/pnnl-preprocessor>

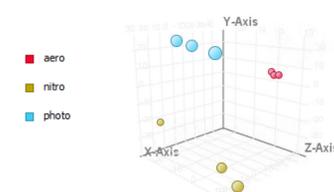
Results

Differential abundance analysis



IM-MS metabolite features

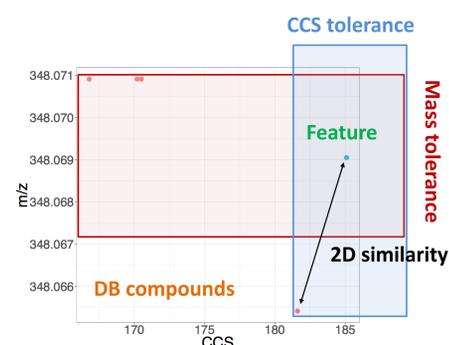
Feature finding and multi-sample alignment performed with Agilent's Mass Profiler. In total 1140 features were detected across the three growth conditions (3 replicates each).



Principal component analysis

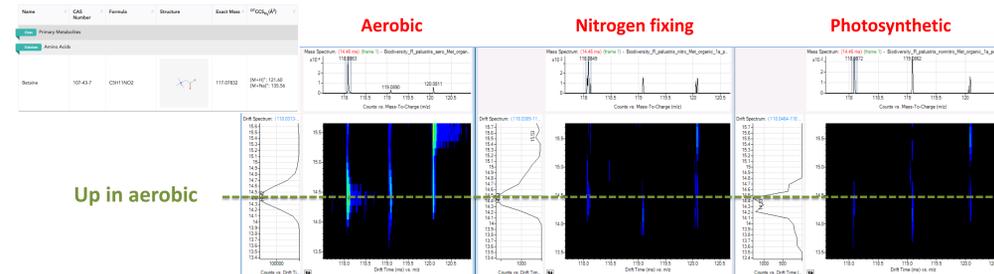
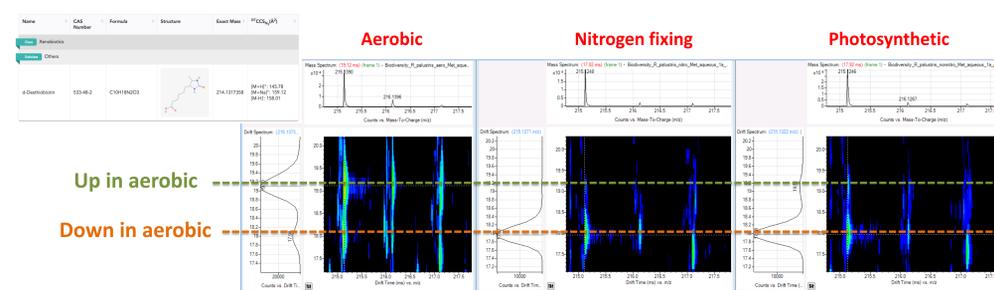
R. palustris metabolite features were imported into Mass Profiler Professional. The three growth conditions are clearly separated along the principal components.

Towards multidimensional scoring for metabolite identification



Example comparing tolerance versus similarity-based scoring

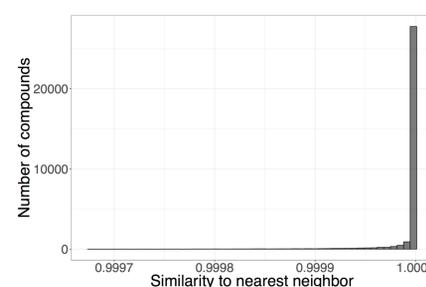
- R. palustris* metabolite feature detected in IM-MS data.
- CCS values predicted by the machine learning-based method MetCCS [3].
- The red box represents 5 ppm tolerance and the blue box represents 1.5% CCS tolerance.
- Using these tolerances, no compound in the database matches both filtering criteria.
- A similarity can combine the 2 dimensions into a multidimensional score and further statistical analysis can be applied to assess the significance of the candidate match.



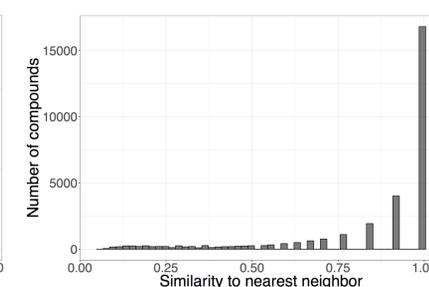
Examples of differentially abundant compounds

Views of the matched compounds from PNNL CCS database [2]. Raw data visualization in IM-MS Browser. Top panels: the peak at later mobility is down-regulated in nitrogen fixing and photosynthetic versus aerobic (matched to the d-Desthiobiotin xenobiotic compound). The peak at earlier mobility is up-regulated in nitrogen fixing and photosynthetic versus aerobic (unknown). Bottom panels: the peak is up-regulated in nitrogen fixing and photosynthetic versus aerobic (matched to the betaine compound).

COSINE SIMILARITY



EUCLIDEAN SIMILARITY



Database similarity analysis

- HMDB compounds (35203), single-charge protonated ions.
- CCS values predicted by the machine learning-based method MetCCS [3].
- For each compound, the similarity to the neighbors within tolerances of 25 ppm mass and 5% CCS was computed and the highest one is compared.
- More than 90% of the nearest neighbors are indistinguishable by cosine similarity. The score is dominated by the mass due to the differences in scale and precision (CCS has only one decimal place and accurate mass has four).
- Euclidean similarity distinguishes about 50% of the nearest neighbors.

Conclusions

- The three different growth conditions of *Rhodospseudomonas palustris* could be distinguished based on the IM-MS results.
- Euclidean similarity provides better discrimination power than cosine but more accuracy and precision are needed in CCS predictions to further distinguish compounds.
- Work in progress to include isotopic pattern and retention time when available in the multidimensional scoring.

Acknowledgments

This research was supported by Agilent Technologies and the Undiagnosed Diseases Network Metabolomics Core grant U01TR001395. Samples were analyzed using capabilities developed under the support of NIH National Institute of General Medical Sciences (8 P41 GM103493-10). This project was performed in the Environmental Molecular Sciences Laboratory, a DOE OBER national scientific user facility on the PNNL campus. PNNL is a multiprogram national laboratory operated by Battelle for the DOE under contract DE-AC05-76RL01830.

References

- Metz TO, Baker ES, Schymanski EL, Renslow RS, Thomas DG, Causon TJ, Webb IK, Hann S, Smith RD, Teeguarden JG. Integrating ion mobility spectrometry into mass spectrometry-based exposome measurements: what can it add and how far can it go?. *Bioanalysis*. 2017 Jan;9(1):81-98.
- Zheng X, Aly NA, Zhou Y, Dupuis KT, Bilbao A, Paurus VL, Orton DJ, Wilson R, Payne SH, Smith RD, Baker ES. A structural examination and collision cross section database for over 500 metabolites and xenobiotics using drift tube ion mobility spectrometry. *Chemical Science*. 2017; 8(11):7724-36.
- Zhou Z, Shen X, Tu J, Zhu ZJ. Large-scale prediction of collision cross-section values for metabolites in ion mobility-mass spectrometry. *Analytical chemistry*. 2016 Nov 1;88(22):11084-91.