Elucidation of Chemical Dark Matter in Soil Using 'Standards-free' Small Molecule Identification

Ryan Renslow[1]* (ryan.renslow@pnnl.gov), Dennis Thomas,[1] Sean Colby,[1] Jamie Nuñez,[1] Yasemin Yesiltepe,[2] Niranjan Govind,[1] John Cort,[1] Mark Borkum,[1] Nancy Washton,[1] Xueyun Zheng,[1] Erin Baker,[1] Justin Teeguarden,[1] Thomas O. Metz,[1] **Kirsten S. Hofmockel,[1] Janet K. Jansson[1]**

[1]Pacific Northwest National Laboratory, Richland, WA; [2]Washington State University, Pullman, WA

https://www.pnnl.gov/biology/programs/

Project Goals: PNNL's Soil Microbiome SFA aims to achieve a systems-level understanding of the soil microbiome's phenotypic response to changing moisture through spatially explicit examination of the molecular and ecological interactions occurring within and between members of microbial consortia. Integrated experiments will be designed to confront both the scaling challenges and inter-kingdom interactions that regulate networks of biochemical reactions. Individual- and population-based models for predicting interspecies and inter-kingdom interactions will be parameterized using experimental data, and predictions will be tested in soil to reveal spatially explicit microbial interactions. Discoveries from controlled experiments will be tested and validated in the field, using moisture gradient experiments at a new local field site. Data will be captured and shared through the establishment of a Soil Microbiome Knowledgebase (SMK). Knowledge gained will provide fundamental understanding of how soil microbes interact to decompose organic carbon and enable prediction of how biochemical reaction networks shift in response to changing moisture regimes.

The ability to unambiguously and comprehensively identify metabolites and other small molecules in complex environmental samples will revolutionize our understanding of metabolic interactions occurring between members of soil microbial communities and between soil microbes and plants. In comparison to genetic information, much less is understood about the identities of small molecules comprising the metabolome, largely due to insufficiencies in molecular identification methods[1]. A significant obstacle in the field of metabolomics is the absence of methods for accurate, rapid, and comprehensive identification of small molecules in complex mixtures without relying on data obtained from analyses of authentic reference materials. This is critical for achieving our goals of identifying metabolic interactions between microbial species and for developing biochemical reaction networks of cellulose and chitin decomposition in grassland soils. We aim to compare the resulting networks to determine consistent or unique pathways across multiple soil sites. Our novel molecular identification pipeline, ISICLE (*In Silico* Chemical Library Engine), uses a large-scale computational chemistry platform that exploits PNNL's high-performance computational quantum chemistry software, NWChem, to calculate metabolite chemical properties, such as collision cross section (CCS)[2,3] and nuclear magnetic resonance (NMR) chemical shift. These properties can subsequently be used to make 'standards-free' identifications of small molecules, required to identify novel metabolic interactions for our project.

In initial tests of our platform, we investigated positional and geometric plant metabolite isomers analyzed using ion mobility spectrometry-mass spectrometry, and found that our platform was significantly more accurate at calculating CCS values compared to other methods, in part due to

Boltzmann weighting of hundreds of candidate conformers by relative energy. This level of accuracy enabled us to even distinguish *cis*/*trans* isomers[4]. Furthermore, we applied ISICLE in calculating CCS for metabolites in the Universal Natural Product Database in order to evaluate the theoretical resolving power of accurate mass and CCS. Finally, we analyzed environmental soil samples and CCSs were calculated *in silico* for possible metabolites. Several predicted degradation products, not available as authentic reference materials, were putatively identified only by accurate mass and *in silico*-derived CCS.

For novel molecule structure elucidation, ISICLE employs density functional theory (DFT) techniques to calculate NMR chemical shifts of molecule sets, with custom options for different solvents, nuclei, and user-selected chemical shift reference compounds. ISICLE calculates NMR chemical shifts of a molecule set with a variety of DFT methods while considering hundreds of conformers for each molecule. NMR chemical shift predictions were validated with experimental data from 300 molecules available in the literature. ^{1}H and ^{13}C chemical shifts were calculated with eight levels of DFT theory, with RMSD errors reaching 0.8 ppm and 5 ppm, respectively. Furthermore, we tested ISICLE on conformers obtained using DFT-based *ab initio* molecular dynamics, demonstrating the ability to reduce chemical shift errors to less than 0.1 ppm (^{1}H) and 2 ppm (^{13}C) using Boltzmann weighting of calculations for hundreds of conformers. Finally, we applied these methods to reassess the identification of wrightiadione as an isoflavonoid extracted from *Wrightia* plants. Based on our calculations, this molecule is actually the alkaloid tryptanthrin, an isobaric isostere of wrightiadione. To our knowledge, the misidentification of wrightiadione has heretofore been unrecognized, and the wrightiadione compound can no longer be said to be known to exist in nature.

Using our platform to identify active and present metabolic interaction is a central aspect of our project. This will allow us to further identify reaction modules (such as chemical and signaling pathways) that are not currently represented in soil biochemical reaction networks.

References

1. Metz, T. O. *et al.* Integrating ion mobility spectrometry into mass spectrometry-based exposome measurements: what can it add and how far can it go? *Bioanalysis* **9**, 81-98, doi:10.4155/bio-2016-0244 (2017).
2. Zhang, X. *et al.* SPE-IMS-MS: An automated platform for sub-sixty second surveillance of endogenous metabolites and xenobiotics in biofluids. *Clinical Mass Spectrometry* **2**, 1-10, doi:https://doi.org/10.1016/j.clinms.2016.11.002 (2016).
3. Zheng, X. *et al.* Enhancing glycan isomer separations with metal ions and positive and negative polarity ion mobility spectrometry-mass spectrometry analyses. *Anal Bioanal Chem* **409**, 467-476, doi:10.1007/s00216-016-9866-4 (2017).
4. Zheng, X. *et al.* Structural elucidation of cis/trans dicaffeoylquinic acid photoisomerization using ion mobility spectrometry-mass spectrometry. *The Journal of Physical Chemistry Letters* **8**, 1381-1388, doi:10.1021/acs.jpclett.6b03015 (2017).

Funding statement: This research was supported by the U.S. Department of Energy (DOE), Office of Biological and Environmental Research (OBER), as part of BER's Genomic Science Program (GSP), and is a contribution of the Pacific Northwest National Laboratory (PNNL) Soil Microbiome Scientific Focus Area "Phenotypic Response of the Soil Microbiome to Environmental Perturbations". A portion of this work was performed in the William R. Wiley Environmental Molecular Sciences Laboratory (EMSL), a national scientific user facility sponsored by OBER and located at PNNL. PNNL is a multi-program national laboratory operated by Battelle for the DOE under Contract DE-AC05-76RLO 1830.