O-35

Session 9) Metabolism & Medical Application of Isotopologue

Pysotopic: A machine learning approach for Chemi-Isotopolomics

Christopher J. Butch[1], Alejandro Lozano Garcia[2], Yufei Miao[1], Shawn McGlynn[3], Johanne Albrigtsen[4], Jennifer Reeve[4], Boswell Wing[4]*

[1] Nanjing University, China

[2] Stanford University, USA

[3] ELSI, Tokyo Tech., Japan

[4] University of Colorado Boulder, USA

*Presenting Author Email: chrisbutch@nju.edu.cn

In this work, we revisit group contribution approaches to the prediction of isotopic fractionation with a modern cheminformatic approach in the hopes of making position specific isotopologue calculations widely accessible. To this end, we present several substantial improvements to the approach first reported by Gallimov. First, we present the largest database of equilibrium isotopic fractionation values reported to date. Collected from both experimental and computational reports, this data contains over 500 molecules representing thousands of contribution motifs for carbon, nitrogen, oxygen, and sulfur. Second, using a ridge regression approach, we calculate bond contribution factors (L-factors) for every motif present in the database, for use in future calculations. Third, we develop a machine learning approach for the prediction of L-factors of bond motifs not present in the database to expand the applicability of our model to broader chemical space. Fourth, we provide a web server for calculations directly from SMILES strings or common molecular file formats for broad ease of use. Throughout our development process we have focused on evergreen development with the hopes that through modularity and transparency the tools and database presented here can serve as community standard well into the future.

The end result of this development is Pysotopic, a tool and database for the rapid calculation of equilibrium stable isotope fractionation at the scale of modern bioinformatics. We achieve a level of position specific accuracy comparable to DFT and MP calculations, in one billionth (10−9) of the time. With this increased speed, comes the ability to investigate previously inaccessible problems such as the position specific fractionation of the Human Metabolome, which we include as supplemental information.