Goldschmidt 2024 Abstract
https://doi.org/10.46247/gold2024.23727

Detection of Agnostic Molecular Biosignatures in Pyrolysis Gas Chromatography-Mass Spectrometry Data using Machine Learning

GRETHE HYSTAD[1], HENDERSON JAMES CLEAVES II[2,3], ANIRUDH PRABHU[4], MICHAEL L WONG[4], COLLIN A GARMON[1] AND ROBERT M HAZEN[4]

[1]Purdue University Northwest
[2]Blue Marble Space Institute for Science
[3]Earth Life Science Institute, Tokyo Institute of Technology
[4]Carnegie Institution for Science

Presenting Author: ghystad@pnw.edu

What is the fundamental difference between the chemistry of life and the chemistry of nonliving systems? We propose that biotic materials have different diversity and distribution of molecules than materials in the nonliving world because biomolecules are selected for their functions through evolutionary processes. We suggest that there exist "agnostic molecular biosignatures" composed of suits of molecules that differentiate terrestrial biochemistry from abiotic chemistry. Even though alien biology might have a different molecular composition, it may still produce interpretable molecular signals via pyrolysis–gas chromatography–mass spectrometry (py–GC–MS) that can be distinguished from those formed through abiotic processes.

We trained machine learning classification models on a diverse collection of natural and synthetic organic molecular mixtures analyzed by py–GC–MS, which resulted in a model that can discriminate between biotic or abiotic samples with ~90% accuracy [1]. The relational aspects of chromatographic retention time, molecular abundance, and mass to charge ratio (m/z) provide the needed information to differentiate between the biotic and abiotic groups. Hence, the method does not rely upon the exact identification of individual compounds.

Using principal component analysis, we find that organic pyrolysis products cluster into three groups: organic material derived from abiotic sources (material from laboratory prebiotic chemistry simulations, carbonaceous meteorites, etc.), living terrestrial matter (living cells, microbes, plants, etc.), and geologically processed biotic organics (including coal, oil shale, petroleum, etc.). We determined the patterns of a diagnostic set of attributes of retention time, abundance, and m/z values that play a role in the observed differences among the abiotic, contemporary biological, and degraded biological specimens. We conclude that these discriminatory features are common in biotic samples and rare in abiotic samples and will discuss how the agnostic nature of our proposed biosignature likely will be useful for the search for evidence of life beyond Earth. The preprocessing steps for the py-GC-MS data will be described where we utilized the full resolution of the 3D structure of the data.

[1] Cleaves et al. (2023), A robust, agnostic molecular biosignature based on machine learning, *PNAS* 140, e2307149120.