Goldschmidt 2023 Abstract
https://doi.org/10.7185/gold2023.18592

A Robust Molecular Biosignature Based on Machine Learning Applied to Three-Dimensional Pyrolysis GCMS Data

ROBERT M HAZEN[1], HENDERSON JAMES CLEAVES[2], GRETHE HYSTAD[3], ANIRUDH PRABHU[1], MICHAEL L WONG[1], GEORGE CODY[1] AND SOPHIA E ECONOMON[4]

[1]Carnegie Institution for Science
[2]Blue Marble Space Institute
[3]Purdue University Northwest
[4]Johns Hopkins University

Presenting Author: rhazen@ciw.edu

The search for definitive biosignatures—unambiguous markers of past or present life—remains a central goal of astrobiology. Our hypothesis is that deeply-rooted aspects of biochemistry differ fundamentally from the chemistry of the nonliving world. Unlike molecules in nonliving systems, life's carbon-based molecular building blocks must be selected for their functions, including to store and replicate information, gather energy and atoms, build structures, control environments, and more. Making functional biomolecules requires energy and information—precious commodities in a competitive Darwinian world. Therefore, we suggest that the diversity and distribution of organic molecules in living systems are different from organic molecular suites produced by abiotic processes, because the biological processes of selection for function and efficient synthesis pathways lead to different frequency distributions of biotic molecules compared to what emerges from purely abiotic processes.

We employed NASA flight-tested pyrolysis gas-chromatography mass-spectrometry (GCMS) methods to analyze 134 varied carbon-rich samples from living cells, taphonomically-degraded samples, geologically processed fossil fuels, C-rich meteorites, and laboratory-synthesized organic compounds and mixtures. Using a suite of machine-learning methods, three-dimensional (time/intensity/mass) data from each abiotic or biotic sample were employed as training or testing subsets, which resulted in a model that can predict the abiotic or biotic nature of the sample with greater than 90 percent accuracy. Furthermore, samples from living cells, geologically-processed biota, and abiotic mixtures reveal discrete attributes that point to the possibility of more granular identification of organic-rich samples (see Figures). Implications include: (1) we can apply this method to Mars and ancient Earth samples to tell if they were once alive; (2) at some deep level biochemistry differs from abiotic organic chemistry; and (3) because of the nature of the method, it is likely that it could distinguish alternative biospheres from that of Earth.

First Figure: Side-by-side pyrolysis GCMS "landscape" averaged views of abiotic (left) and biotic samples reveal distinct topologies.

Second Figure: Two (of 20) principal components illustrate discrimination between biotic (red) and abiotic (blue) samples. Furthermore, geologically-processed biotic samples (e.g., coal; petroleum) follow a trend distinct from living cells, suggesting a more granular discrimination is possible.