Goldschmidt 2026 Abstract
https://doi.org/10.46427/gold2026.36626

Uncovering organic geochemical evidence for life and photosynthesis in Archean rocks via pyrolysis–GC–MS and supervised machine learning

MICHAEL L WONG[1], ANIRUDH PRABHU[1], EMERSYN SLAUGHTER[2], EMILY P. DOBBERFUHL[3], GEORGE D. CODY[4], GRETHE HYSTAD[5], HENDERSON JAMES CLEAVES II[6] AND ROBERT M HAZEN[1]

[1]Carnegie Institution for Science
[2]University of Florida
[3]University of Wisconsin, Madison
[4]Earth and Planets Laboratory, Carnegie Institution for Science
[5]Purdue University Northwest
[6]Howard University

Teasing out biochemical information from ancient organic-rich sediments—including robust molecular evidence for microbial activity, as well as the timing of the emergence of photosynthesis relative to the inferred oxygenation of Earth's atmosphere—remains a challenging opportunity. To tackle this problem, we analyzed hundreds of diverse ancient and modern samples using pyrolysis–gas chromatography–mass spectrometry (py–GC–MS), then trained a supervised machine learning (ML) model on multidimensional patterns of molecular data to discriminate samples of biogenic vs. abiogenic origin, as well as photosynthetic vs. nonphotosynthetic physiology. Comparing organic-rich samples of uncertain affinity to our training data, ca. 3.33-billion-year-old sedimentary rocks group among microbial samples, and rocks as old as 2.52 billion years show signs of phototrophy [1]. Such findings effectively double the interval within which fossil organic matter can be shown to retain molecular information of evolutionary relationships and physiology.

Ongoing research investigates our ML-based molecular biosignature technique's relationship to traditional biomarkers and its quantifiable, probabilistic assessment of heterogeneous samples. By employing "interpretable" machine learning algorithms, like random forest classification, we can identify the specific features responsible for biotic vs. abiotic discrimination [2]. Current work demonstrates that our method can detect signs of life even when biogenic matter is mixed with organics of abiotic origin [3] and correctly classifies ancient biotic matter that has been contaminated with small amounts (<10 wt%) of modern biological material.

Furthermore, our technique can be extended to the realm of astrobiological exploration [4], where it can be used to identify "potentially biological anomalies," i.e., possible signs of life with non-Earthly biochemistry. Similar ML approaches may be applicable to other kinds of instruments, or to create a unified framework for biosignature assessment that considers data from multiple analytical instruments at once. Overall, our technique has the potential to serve the goal of life detection and characterization across both space and time.

[1] Wong et al. (2025) *PNAS*, 122, 47, e2514534122
[2] Hystad et al. (2025) *JGR Machine Learning & Computation*, 2, e2024JH000441
[3] Slaughter et al. (in review) *Frontiers in Astronomy & Space Sciences*
[4] Cleaves et al. (2023) *PNAS*, 120, 41, e2307149120