Goldschmidt 2026 Abstract
https://doi.org/10.46427/gold2026.36133

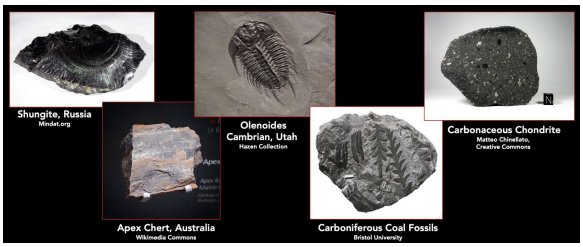
A machine learning approach to molecular biosignatures: A new strategy in the astrobiological search for life

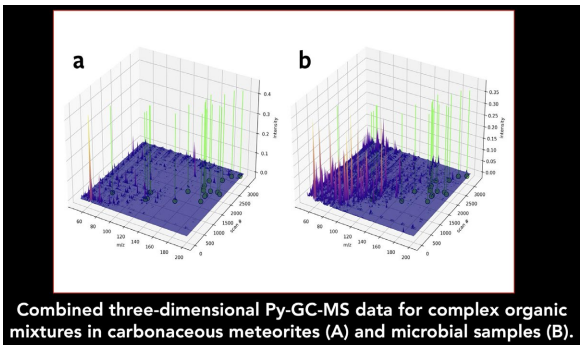
ROBERT M HAZEN, MICHAEL L WONG AND ANIRUDH PRABHU

Carnegie Institution for Science

Throughout Earth's history, organic molecules from both abiogenic and biogenic sources have been buried in sedimentary rocks. Most of these organic molecules have been significantly altered by geologic processes through deep time. Nonetheless, the nature and distribution of those ancient fragmentary organic remains have the potential to reveal diagnostic biomolecular information after billions of years of burial. We analyzed 406 fossil, modern biological, meteoritic, and synthetic samples [Figure 1] using pyrolysis gas chromatography and mass spectrometry [1]. We explored these analytical data via supervised machine learning methods to discriminate samples of biogenic versus abiogenic origin, plant versus animal phylogenetic affinity, and photosynthetic versus non-photosynthetic physiology [Figure 2]. Dividing 272 samples with known phylogenetic affinity and physiology into 9 categories, each further divided into 75% training and 25% testing sets, our Random Forest models accurately predict pairwise assignments of modern versus fossil or meteoritic organics (100% correct assignments), fossil plant tissues versus meteoritic organics (97%), modern versus fossil plant tissues (98%), and modern plants versus animal tissues (95%). Pairwise comparisons between fossil biogenic samples versus abiogenic samples resulted in 93% correct classifications, while analysis of modern and ancient photosynthetic versus non-photosynthetic samples also resulted in 93% correct assignments. Our analyses demonstrate that molecular biosignatures can survive in ancient fossils and allow for the identification of organismal origins and traits. Consistent with previous morphological and isotopic inferences, we present evidence for biogenic molecular assemblages in Paleoarchean rocks (3.33 Ga) and for photoautotrophy in Neoarchean rocks (2.52 Ga). Work in progress will merge pyrolysis-GC-MS data with those of other analytical methods, including Raman and IR spectroscopy, stable isotope analyses, and other attributes to develop a multi-messenger, flight-ready life detection system.

[1] M.L. Wong, A. Prabhu, et al. (2025) *PNAS* 122, e2514534122.





Combined three-dimensional Py-GC-MS data for complex organic mixtures in carbonaceous meteorites (A) and microbial samples (B).