Goldschmidt 2026 Abstract
https://doi.org/10.46427/gold2026.36291

Benchmarking In Silico Mass Spectral Libraries for Characterising Canadian Oil Sands Naphthenic Acids

DR. CHARLES E. WEST, PHD[1], STEVEN ROWLAND[2], MARK HEWITT[3] AND RICHARD FRANK[3]

[1]Independent
[2]University of Plymouth
[3]Environment Canada

Naphthenic Acids (NAs) are ubiquitous components of organic matter in diverse geochemical environments, making their characterization vital for better understanding petroleum formation, biodegradation and the environmental legacy of industrial processes. While NAs are found across many natural systems, they are of particular concern in oil sands process water (OSPW), where structural identification is important for source fingerprinting and risk assessment. Due to the extreme isomeric complexity of these mixtures and the limited availibility of synthetic reference compounds, *in silico* mass spectral libraries, such as those generated by GCMS-ID (incorporating CFM-ID 4.0) have emerged as a potential resource for suspect screening. However, the reliability of these machine-learning models may not be uniform across structural classes, a discrepancy that justifies further investigation.

In this study, we benchmarked *in silico* predicted electron ionization (EI) mass spectra against experimental spectra from a library of novel, synthetic tricyclic NA methyl esters (ME) representing two distinct topologies: caged (adamantane type) and non-caged (fused-ring) monoaromatic architectures; both key structural types identified in OSPW fractions that present significant analytical challenges. Experimental spectra (EI-MS) obtained for the synthetic reference compounds were compared to *in silico* predictions using National Institute of Standards and Technology (NIST) match factors.

Our results reveal predicted EI mass spectra for rigid, caged adamantane acids (ME) achieved moderate match scores of 700-760, whilst those for fused-ring monoaromatic tricyclic isomers consistently yielded poor match scores (500-690). This marked drop in prediction capability for fused ring architectures compared to caged type structures, highlights a critical structural dependency in current algorithmic performance. Rather than precluding the use of *in silico* tools, these findings demonstrate a clear pathway for enhancement: the synthesis and characterization of diverse, high purity reference standards* is essential to expand training sets and improve the robustness of predictive models. By incorporating these complex but justified structures, the geochemical community may overcome current limitations in automated predictive models, ultimately enabling more definitive characterizations of NAs across complex natural systems, ranging from biodegraded petroleum and oil sands assessment to OSPW environmental monitoring.

*We acknowledge the efforts of the late Professor Victor Snieckus and colleagues at Snieckus Innovations.