# IDENTIFYING UNKNOWN SPECTRAL LINES IN JWST EXOPLANET SPECTRA

LAURA K McKEMMISH, MARIA PETTYJOHN, JUAN C. ZAPATA TRUJILLO, RHYS McALISTER, NATALIA ADAMOWSKA, *School of Chemistry, University of New South Wales, Sydney, NSW, Australia*.

The James Webb Space Telescope has transformed our understanding of exoplanet atmospheres, revealing unidentified spectral lines whose molecular origins could challenge existing models of exoplanet chemistry, physics, geology, and potentially biology. Our role as spectroscopists is to help astronomers decipher these unexpected signals. To achieve secure detections, constructing high-accuracy line lists for specific molecules at elevated temperatures typically requires years of experimental and theoretical effort. However, an essential initial step is to shortlist candidate molecules responsible for these unknown lines.

Our innovative high-throughput computational approach provides approximate fundamental vibrational band centers and intensities, along with error estimates, for nearly 2000 molecules containing up to 6 non-hydrogen atoms. This methodology can be extended to generate data for any plausible non-metal gaseous molecule. We are also developing techniques for approximate cross-sections at different temperatures using rotational constants, anharmonic approaches for overtone and combination bands, and methods to model vibrational hot bands. Benchmarking is crucial in all theoretical approaches to provide reasonable error estimates.

With our dataset of approximate spectral data with estimated errors, we can identify a shortlist of molecular candidates for each unknown band. This list can be evaluated through manual literature searches for additional experimental data and targeted, very high accuracy quantum chemistry calculations. Our initial tests show that many candidate molecules are chemically unconventional and unlikely to be present. Incorporating a more data-driven and semi-automated estimation of molecular abundance in certain atmosphere classes would enhance our approach.

The dataset's size prompts consideration of using machine learning to explore a larger set of molecules, particularly bigger ones relevant in cooler gaseous environments like solar system moons. I will briefly present preliminary results highlighting the importance of data quality in ensuring the accuracy of machine learning models. Model evaluation should be based on experimental data rather than computationally-generated training data. Molecular similarity analysis can also assist in predicting the uncertainty of predictions for new molecules.