# Automated Discovery of Important Chemical Reactions

Colin A. Grambow

## Technical Summary

Innovations in chemistry are often informed by decades of accumulated chemical knowledge encoded into manually constructed reaction templates and rules of reactivity. Examples include retrosynthetic analysis for organic synthesis planning; chemical reaction mechanism generation for complex combustion, pyrolysis, and low-temperature oxidation processes; and elucidation of low-energy catalytic pathways. Nonetheless, all known chemistry is dwarfed by the vastness of chemical space, most of which still lies unexplored. *De novo* reaction discovery is rare but presents an enormous potential to uncover novel synthetic routes and key pathways in reaction mechanisms. Automated potential energy surface exploration has become a promising method to search for new reaction pathways, albeit at the expense of costly quantum mechanical calculations.

Therefore, this thesis develops methods to enable more computationally efficient discovery while also correctly determining thermochemistry and kinetics to allow for the construction of accurate reaction mechanisms.

By utilizing automated transition state finding algorithms based on quantum chemistry, the thesis assesses which algorithm is most viable for the efficient discovery of new reactions, and it identifies key pathways of an important ketohydroperoxide system. It demonstrates that quantum chemical data can be used with emerging machine learning methods to estimate molecular thermochemistry. Leveraging a large data set of low-quality data in combination with a small data set of high-accuracy data in a transfer learning approach enables predictions that significantly improve upon group additivity methods, which are common in automated mechanism generation, and upon machine learning models that only use density functional theory data. Furthermore, an automated workflow is developed to further enhance high-level quantum chemistry calculations using bond additivity corrections.

While quantum chemistry calculations are incredibly useful at providing highly accurate data, their high cost—especially when applied to thousands of reaction pathways—limits their utility for discovering new chemistry. Therefore, this thesis improves the throughput of automated discovery via a combination of quantum chemistry data generation and reactivity prediction using deep learning. It automatically generates a data set of tens of thousands of elementary chemical reactions that are used to train a novel activation energy prediction model, which can quickly assess the importance of new reactions.