Automated Chemical Space Exploration Using Machine-Learned Potential and Data-Driven Property Prediction

The exploration of chemical space, particularly the local configurational space of molecular clusters, remains a significant challenge due to the combinatorial growth of possible structures, rendering traditional brute-force approaches computationally infeasible. This thesis introduces a machine learning–enabled growth strategy that significantly reduces the cost of exploring such configurational landscapes. We develop a modular bottom-up cluster-generation approach leveraging the transferable AIMNET2 potential, which provides DFT-level trends for energies and geometries at significantly reduced computational cost, with demonstrated transferability within the studied chemical domains.

Furthermore, this framework is extended to hydrocarbon exploration, where we construct an expanded hydrocarbon database up to C_{10}, enabling the identification of high-energy candidates relevant to aviation fuels and interstellar chemistry. This screening yields 23 strained hydrocarbon candidates, of which 20 are not reported in existing databases to the best of our knowledge, and these are cataloged in the **Hydromol** web application. Finally, we apply a donor–acceptor (D–A) building-block strategy to explore the local chemical space of thermally activated delayed fluorescence (TADF) molecules. Using a workflow combining AIMNET2, CREST, and multi-layer TD-DFT calculations, we generate a database of 30,987 compounds, from which 42 promising TADF candidates are identified and 8 representative molecules are analysed in detail.

This work emphasises efficient exploration of local configurational landscapes over exhaustive enumeration of global chemical space, offering a practical and scalable framework for accelerated molecular discovery.

Keywords: Chemical space exploration, Molecular clusters, Machine learning potentials, AIMNET2, Global optimisation, TADF materials.