

FROM MOLECULES TO POLYMERS AND BEYOND: MACHINE LEARNING-AIDED MATERIAL STABILITY AND REACTION KINETICS PREDICTION

–Veerupaksh Singla

This work presents a comprehensive framework for chemical stability and reaction prediction across diverse material classes, using progressively sophisticated computational approaches. It addresses a critical gap in materials design: the lack of efficient methods for predicting degradation susceptibility, which has long hindered innovation in high-performance materials. Through a series of interconnected studies, this research demonstrates how machine learning can transform reaction kinetics data into actionable stability predictions, providing a scalable pathway from simple molecules to complex materials.

The research began by establishing that chemical stability is a fundamentally learnable scalar metric derived from computational kinetic data. Using the thermal degradation of alkanes as a model system, half-lives of approximately 32,000 acyclic alkanes under pyrolysis conditions were simulated based on literature-derived radical elementary steps and graphically constructed model reactions. Machine learning models, trained on these simulations with a hinge-loss function, could predict relative stability directly from molecular graphs. The models successfully captured known trends related to branching and molecular size, demonstrating remarkable transferability across alkane subclasses. This initial study showed that even relatively simple machine learning models could effectively learn complex stability relationships, laying the groundwork for broader applications.

Building on this foundation, it was hypothesized that small-molecule kinetics could predict polymer thermal stability without polymer-specific training data, given the local nature of degradation initiation. Using only small-molecule kinetic data, the models were able to predict the relative thermal stability rankings of 41 common polymers containing C, H, O, N, F, and Cl. Remarkably, models trained exclusively on small molecules showed strong transferability to polymers. A key insight emerged: while chemical diversity generally improved model performance, the consistency and quality of the underlying kinetic data were equally critical. Models trained on high-quality alkane data performed comparably or better than those trained on more diverse but inconsistent datasets, highlighting that accurate small-molecule kinetics can effectively proxy for polymer stability—opening new avenues for early-stage material design.

The success of these predictions underscored the need for more accurate kinetic data across a broader chemical space. Recognizing that machine learning models often struggle to extrapolate beyond their training domain, a comprehensive dataset of approximately 12,000 DFT-calculated reactions, derived from

around 450 graphical model reactions, was developed. This dataset, containing transition state geometries and thermochemistry, was designed to streamline reactive pattern detection and improve model generalizability. Benchmarking confirmed that model reaction kinetics could serve as reliable surrogates for actual reaction kinetics, easing machine learning tasks. The dataset and associated code were made publicly available, addressing a major bottleneck in computational reaction prediction.

This research culminated in the development of highly accurate methods for predicting reaction activation barriers scalable to larger systems. Δ -learning models were trained to correct errors from graphical reaction templates and predict DFT-level activation barriers for reactions involving up to 22 heavy atoms. Two approaches were employed: XGBoost with differential reaction fingerprints (DRFPs) and ChemProp, a message-passing graph neural network. Δ -learning consistently improved barrier prediction accuracy by about 1 kcal/mol for interpolation tasks and achieved chemical accuracy (<1 kcal/mol MAE) in external tests on unradical alkane pyrolysis reactions. Moreover, targeted transfer learning—using single-point calculations at higher theory levels on just 14% of the dataset—further reduced errors by 30–50% for challenging systems. Together, these results establish Δ -learning as a scalable, accurate, and computationally efficient framework for reaction barrier prediction.

In sum, this thesis presents an integrated approach to predicting material stability, progressing from fundamental principles to practical applications. By demonstrating that stability can be learned from kinetic data, validating transferability from small molecules to polymers, building expansive reaction datasets, and developing advanced Δ -learning models, a robust framework for predicting stability across diverse material classes has been established. This work addresses key challenges in computational materials science by balancing accuracy, transferability, and computational efficiency, offering powerful tools to accelerate the discovery and design of stable, high-performance materials. Future directions include extending these methods to more complex material systems and degradation mechanisms, further enhancing our ability to design materials with tailored stability profiles.