Property prediction with machine learning and ab initio methods for iridium photoactive complexes and metal-organic frameworks

by
Gianmarco Terrones

Data-driven prediction of a chemical's properties prior to synthesis or use can accelerate chemical discovery by increasing the probability of candidate suitability for the given application. In this thesis, data-driven models and complementary first-principles calculations have been used to study three types of chemistries: iridium photoactive complexes, metal-organic frameworks, and reactions for azetidine synthesis. Iridium photoactive complexes are commonly used in OLED lighting, photocatalysis, and bioimaging due to their unique phosphorescent properties and triplet excited state population. Metal-organic frameworks are studied for heterogeneous catalysis and gas separations and storage due to their tunable metal environments and porous structures. Azetidine-containing molecules have potential for use as pharmaceuticals due to their high stability and good pharmacokinetics. However, despite their advantageous properties, challenges remain in designing these chemistries and informing this design with computation. The excited state properties of iridium complexes are challenging and costly to predict through first-principles methods. Similarly, stability issues often affect metal-organic frameworks, yet these cannot be efficiently modeled by physics-based routines. Lastly, synthetic approaches toward azetidine synthesis are limited, and computational study of novel synthetic approaches to identify desirable reactant characteristics would benefit the future scope of azetidine products.

The models developed in this thesis have proven to be complementary tools to first-principles approaches, and have major benefits in their speed of application and ability to train directly on experimental data for properties that challenge methods like DFT. The models are applied to screen chemical space for promising candidates through consideration of hypothetical, not-yet-synthesized iridium complexes and metal-organic frameworks generated through component combination of existing structures. The machine learning models are also used to derive structure-property relationships through feature importance analysis, for example identifying qualities of iridium photoactive complexes that impart longer or shorter excited state lifetime. In addition to model generation, work in this thesis has covered code development of a software package for molecule structure generation, modification, and fingerprinting, and also development of intuitive web interfaces for easy use of data-driven models. It is expected that the tools developed in this thesis will both allow for a greater understanding of iridium complexes, metal-organic frameworks, and azetidine synthesis, and enable low-cost exploration of chemical space for novel material selection.

Thesis Supervisor: Heather J. Kulik
Title: Lammot du Pont Professor of Chemical Engineering