

Models and Methods for Efficient and Accurate Transition Metal Catalyst Discovery with Density Functional Theory

Daniel B.K. Chu

Technical Summary

Transition metal complexes (TMCs) present great opportunities, and great challenges, for applications in chemistry, such as the catalysis of greenhouse gases into value-added products. TMCs are highly tunable via the metal identity, metal oxidation state, and ligand coordination environment. A great number of chemistries are accessible through this tunability, combined with the flexibility of TMCs over solid-state metal catalysts, make TMCs quite attractive to perform catalysis that cannot be done on solid-state metals. However, it is impossible to explore exhaustively the sheer number of possibilities in ligand coordination environments, let alone the metal identity and oxidation state. Furthermore, the accuracy of quantum chemical calculations, such as density functional theory (DFT), is more variable for metal-containing systems. Thus, it is necessary to devise methods for the efficient and accurate exploration of the TMC design space.

This thesis addresses these issues in part by introducing models which relate the properties of TMCs to each other based on their ligand coordination environments. These models use a ligand many-body expansion to relate properties of low-symmetry TMCs to those of higher symmetry TMCs. While chemical accuracy for transition metal chemistry is around 3 kcal/mol, the *cis* model can infer DFT reaction energetics within 1 kcal/mol on average while only considering the interactions of ligands which are *cis* to each other (nearest-neighbor). Finally, in a Δ -learning type approach, the *cis* model can infer the difference between DFT and more accurate wavefunction theory (WFT) calculations with one-third as much error as with inferring the WFT result alone.

When investigating N monodentate ligands, the introduced models require $O(N^2)$ parameters, leading to an N^4 speedup in the investigation of octahedral TMCs and an N^3 speedup for square-pyramidal TMCs. However, this can still be too expensive to perform calculations for all parameters in the *cis* model. For example, with $N=500$ monodentate ligands there are nearly 4 trillion square-pyramidal complexes, and the *cis* model requires over 100,000 parameters (or measurements). An additive model (i.e., a truncated form of the *cis* model) requires only 999 parameters and can be used as an initial pass to determine which ligands are worthwhile of further investigation. Ultimately, this allows for exploration of a design space of nearly 4 trillion TMCs with calculations or measurements for under 2,000 of the TMCs.

Thesis supervisor: Prof. Heather J. Kulik
Title: Lamot du Pont Professor of Chemical Engineering