Developing Predictive Tools for Solvent Effects on Thermodynamics and Kinetics

Yunsie Chung

January 9, 2023

Solvents are ubiquitous in many industrially, environmentally, and medically relevant chemical systems. Solvents can strongly affect species thermochemistry and reaction kinetics, and a different solvent choice can lead to completely different reaction outcomes and phase equilibria. Accurate prediction of solvent effects is thus crucial to the design and optimization of chemical processes and the construction of liquid phase kinetic models. Ab initio methods such as quantum chemistry can be used to compute solvation effects, but a high computational cost makes them unsuitable for large-scale applications. Furthermore, existing methods are largely limited to the predictions at room temperature. Data-driven approaches like group contribution and machine learning can provide fast estimations, but a lack of quality data is a major bottleneck to these approaches.

This thesis presents several new models that can provide fast and accurate predictions of solvent effects on thermodynamics and kinetics for a wide range of chemical space and temperature. The approaches employed in this work center around combining fundamental thermodynamic relationships, quantum chemistry, and machine learning. An extensive set of quantum chemical data is generated with ab initio methods and used to train machine learning models. Thermodynamic equations and correlations are used to make predictions for different properties and conditions based on available or calculable data. The devised models and methods can provide accurate estimates of temperature-dependent solvation free energy, solvation enthalpy, and solid solubility. The predictions can be made up to the critical point of a solvent, allowing one to simulate gas-liquid and solid-liquid equilibria for the entire range of temperature. Various quantum chemistry and COSMO-RS levels of theory are compared to identify an efficient and reliable computational workflow for the calculation of liquid phase rate constants. After establishing the optimal workflow, large-scale COSMO-RS calculations are performed and a machine learning model to predict kinetic solvent effects on neutral reactions is developed. The performances of all models are thoroughly evaluated by a direct comparison with the experimental data that are compiled from numerous public sources. The presented tools only need molecular identifiers or easily obtainable data as inputs and hence are ideal for automated, high-throughput applications.

Thesis Supervisor: William H. Green
Title: Hoyt C. Hottel Professor in Chemical Engineering