Constructing a Cyberinfrastructure for Automated Reaction Mechanism Generation

Michael Harper
Thesis Advisor: William H. Green

Over the past decade, members of the Green Group have been constructing a software package known as Reaction Mechanism Generator, or RMG (http://sourceforge.net/projects/rmg/). Using reaction family templates, e.g., “Isomerization,” “H-Abstraction,” and “Diels-Alder,” RMG can automatically generate a reaction mechanism – a list of species, reactions, and the associated Arrhenius parameters – for any reactant(s) (containing only carbon, hydrogen, and oxygen atoms), given a set of initial conditions, e.g., temperature, pressure, and desired conversion.

To generate the reaction mechanism and to predict species concentrations as a function of time, the software requires reaction rate coefficients and thermodynamic parameters for all reactions and species generated in the model, respectively. Ideally, the software would utilize the scientific community’s recommended reaction rate coefficients and thermodynamic parameters. However, experimental data or even quantum chemical calculations are typically unavailable for most species and reactions. Thus, RMG currently relies predominantly on Benson’s group additivity rules to estimate thermochemical data not stored in its database. While Benson’s group additivity theory is well-established, it is not 100% reliable. In an effort to utilize more accurate thermochemical parameters, we are working toward enabling RMG to search the PrIMe Warehouse (http://www.primekinetics.org/), a community-collaborated database, for the scientific community’s recommended numbers before applying empirical estimations.

The union of RMG and the PrIMe Warehouse results in the titled cyberinfrastructure. The first step in constructing the cyberinfrastructure was programming RMG to talk with the PrIMe Warehouse and vice versa. To facilitate the use of chemical databases by computer software without human intervention, it is paramount that the databases have unique chemical identifiers. RMG utilizes an “adjacency list” to identify species whereas PrIMe’s warehouse uses a “primeID.” To bridge the gap between RMG and PrIMe, a unique species identifier, the IUPAC International Chemical Identifier (InChI), was employed. To date, approximately 5,400 species in PrIMe’s species warehouse have been populated with InChI identifiers, which accounts for approximately half of PrIMe’s total species warehouse. Using the InChI string as search criteria, several hundred duplicates were identified and sequentially removed from the database. The PrIMe species catalog now consists of a “clean” C/H/N/O database, that is, one without duplicate species and with unambiguous unique identifiers for each molecule.

We have employed this cyberinfrastructure to generate a detailed reaction mechanism for the combustion of 1-butanol. The mechanism has been validated by comparing the model’s predicted concentration profiles with existing experimental data from (1) pyrolysis of 1-butanol in a plug flow reactor and (2) a methane flame doped with 3500ppm 1-butanol.