A Systematic Approach to Predicting Combustion Chemistry

William H. Green, Department of Chemical Engineering, Massachusetts Institute of Technology, 77 Massachusetts Ave., Rm. 66-270, Cambridge, MA 01239

Overview

To accurately predict the combustion chemistry of liquid or solid fuels we will need a new approach. First of all, the many assumptions and approximations made when constructing the large chemical kinetic models required need to be documented in a way that is human-readable, so they can be easily be peer-reviewed and checked. Second, the process of moving from these assumptions and approximations to a large computer simulation must be completely automated, using well-understood and checked procedures, so that different groups making the same approximations and assumptions will obtain the same numerical results. Third, the process of comparing model predictions with experimental data needs to be significantly upgraded, so that one can unambiguously determine whether or not a particular set of experimental measurements is consistent with a combustion model.

In this talk, a new data model for chemical kinetics is presented, which clarifies the chemistry knowledge and assumptions that underlie the large chemical kinetic models used in combustion. With the new data model, it is feasible to remove most of the chemistry information and hidden assumptions typically hard-coded in automated mechanism-generation software, and bring them out into a human-readable database. This has significant practical advantages, including simplified maintenance of the software as chemistry knowledge improves. In the new data model, chemistry information is organized along intuitive functional-group lines. The new data model should facilitate peer-review and comparisons between even very large combustion chemistry models, and clarify where there are real differences in what is assumed about the chemistry. New model-construction, model-reduction, and least-squares fitting algorithms are presented. Applications relevant to design of new engines and fuels will be presented.

I. Introduction

I.A. The Goal of Kinetic Modeling

The primary goal of chemical kinetic modeling is to make predictions: given our current understanding of chemistry, what do we expect to happen in a particular reacting mixture under specified reaction conditions? This is quite a challenging task, both because “Chemistry Knowledge” is very large and somewhat amorphous, and because it often requires hundreds of differential equations and thousands of numerica parameters (rate constants, molecular thermochemistry, etc.) to accurately describe the details of the processes occurring in a reacting mixture.

Because the “Chemistry Knowledge” input, the procedure for developing the correct list of differential equations corresponding to this input, and the estimates of the numerical parameters are all known imperfectly, it is quite important that each aspect be carefully documented to facilitate scientific progress. Ideally, all the assumptions and simplifications made at each of the many steps from “Chemistry Knowledge” through to quantitative predictions should be clearly and precisely documented, so each can be tested, and each number in both the inputs and the outputs should be assigned an uncertainty band. Clarity about the assumptions and simplifications facilitates identification of the root causes why different models give different predictions, and why predictions differ from experimental results. Clear understanding and documentation of the origins of these discrepancies will dramatically improve the efficiency of the world-wide efforts to improve both chemical understanding and kinetic modeling methodology.

There is more than just scientific progress at stake. Predictions based on chemical kinetic models are increasingly used to inform major policy and business decisions, often with large impacts on society, so it is critical that the uncertainties and assumptions associated with these predictions be clearly enunciated and understood. Clarity about our current level of ignorance is essential both to avoid misleading decision makers and to facilitate future work to test the assumptions and reduce the uncertainties in the predictions.

I.B. Data Models for Chemical Kinetics

Clarity in expressing information (including information about assumptions and uncertainties) can be greatly facilitated by using a good “data model”. A “data model” is a standard way of expressing information; it includes both the format and the relationships between various data objects.

Currently, the most popular way to document a chemical kinetic model is to list the reactions and corresponding rate parameters (typically in a CHEMKIN® format [Kee et al. (1989)]), and separately to list the molecular parameters such as enthalpies and heat capacities (typically in a NASA polynomial format). This popular data model, Fig. 1, has many positive aspects: it unambiguously expresses all the information necessary to construct and solve the differential equations, and it is flexible enough that it can be used for a broad range of chemical systems.

However, the popular List-of-Reactions data model has several serious deficiencies. First of all, it is difficult in this data model to adequately document the origins of all the numerical parameters used in the simulation, and the associated uncertainties in these parameters. The format is also not very user-friendly: it is almost impossible for a human to check that all the numbers and reactions in a large kinetic model expressed in this format are consistent and reasonable, much less correct. This data model also does not allow for graphical representation of chemical structures, leading to sometimes serious inconsistencies in naming conventions.

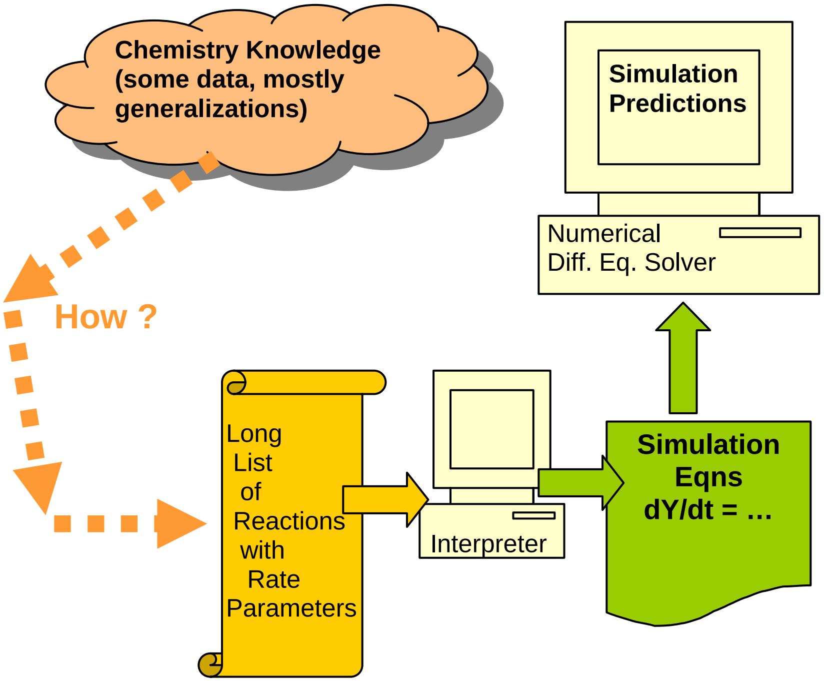

Fig. 1. In the most popular approach, the primary documentation of a kinetic model is the List of Reactions, and the steps from that List to the predictions are often well-documented. However, the steps upstream of the List of Reactions are usually poorly documented.

More fundamentally, the current data model does not provide any way to document why some chemical reactions were included in the simulation, while others were left out. Typically, a large number of assumptions about which reactions/species are likely to be important (under the reaction conditions of interest to the simulation's author) are made by the person (or computer program) who assembles the list of species and reactions, based on his or her (or its) chemistry knowledge. Because the current data model does not provide any convenient way to document these assumptions, most of these assumptions are never documented at all, and this contributes to the impression that this information is not worth recording. The fact that it is usually very difficult to uncover complete information about the assumptions behind a reaction list discourages careful review of the critical steps "upstream" of the CHEMKIN-format file.

It is important to realize that the CHEMKIN-format file really represents an intermediate step in the process of predicting chemical behavior. For large mechanisms, almost all of the parameter values in this data file are estimates that come from some simple (but not always documented, or consistently applied) rate-estimation rule. Similarly, almost all of the numbers in the thermo files are estimates and theoretical extrapolations; in large models the thousands of NASA coefficients in the thermo file are typically derived from a much small set of Benson-type [Benson (1976)] group additivity values. So although the typical CHEMKIN-type file contains thousands of numbers, the real information content is usually much smaller: the numerical values in the simulation could usually be represented much more compactly in terms of group values, and rate-estimation parameters. Reducing the number of numerical parameters would greatly facilitate the important (but now rarely performed) task of comparing two competing kinetic models, to understand why they give different predictions.

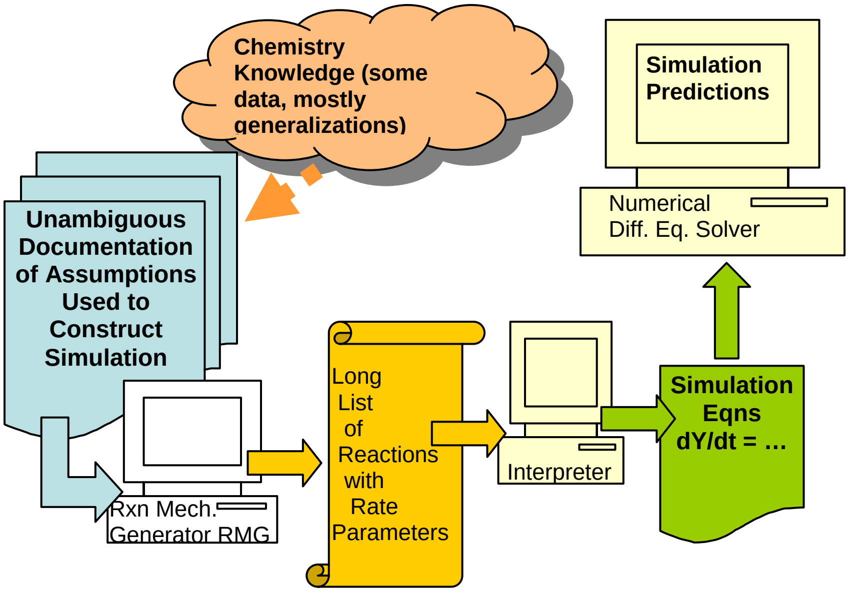

Fig. 2 In the new data model, the fundamental documentation is the information on how one estimates reaction rates (and molecular properties). All the subsequent steps in the process are automated, well-documented procedures

Here we propose a new data model, Fig. 2, where the crucial documentation lies upstream of the conventional list-of-reactions. In this new data model, the estimation rules for the rates and thermochemistry are the main data, supplemented by a set of numerical parameters for individual species and reactions (whose accurate values are known from experiments or high quality quantum calculations.) This new data model overcomes many of the objections to the List-of-Reactions data model it replaces. However, for this data model to be successful, it is necessary that unambiguous procedures, with clearly documented assumptions and tolerances, exist for converting the Estimation Rules into the corresponding List-of-Reactions, and thence (as is done currently) to Predictions. It is most convenient if these procedures can be performed automatically, using a computer.

I.C. Automated Construction of Lists-of-Reactions

The awkwardness of the List-of-Reactions data model became apparent to many researchers modeling pyrolysis and combustion soon after CHEMKIN was developed, and since the 1980's many software packages have been developed to automate the process of constructing the long lists of reactions required, as reviewed by Tomlin *et al.*(1997). Many calculations have been published demonstrating that automated mechanism-construction can be successful. However, all of these software packages had significant flaws, many have been abandoned, and to our knowledge none of this software has yet been successfully distributed by its author to another research group.

To understand the problem, it is helpful to think about what the inputs are to these computer programs, i.e. what does one need to know in order to construct the appropriate List-of-Reactions chemistry model, and so correctly predict the behavior of a reacting system?

I.D. The Fundamental Inputs to a Chemical Kinetic Model

When one begins to construct a chemical kinetic model, there are several different types of required “input” information. Obviously, one needs some specification of the initial concentrations of the reactants, and of the reaction conditions (e.g. T, P, timescale) of interest. Normally one wants to numerically solve the kinetic model to predict species and/or temperature profiles, so the inputs must also include some specification of numerical tolerances on these outputs, and options for the differential equation solver. The most complicated “input” information required to construct a kinetic model is the chemistry: what species, reactions or reaction types will be considered? How will all the thermochemical and rate parameters be estimated?

A naïve answer is that one should just assemble the list of all known species and all the known elementary-step reactions connecting them in the literature, and use all the literature values for the rate and thermochemical parameters. However, except in a very small number of very simple cases (e.g. H_2/O_2 combustion) this naïve approach is both inefficient and seriously inaccurate. First of all, unselectively constructing a kinetic model out of a very large set of reactions has the extremely undesirable effect of making it difficult or impossible for a human to understand / check / peer-review the model. Also, almost invariably, most of the reactions in any large compilation are unimportant under the specific reaction conditions of interest, so one is doing a lot more work both in checking and in solving the large model than is necessary. But the biggest problem is that in most branches of chemistry, only a very small fraction of the important elementary-step reactions have ever been studied. So models constructed by just assembling a list of previously studied reactions from the literature are almost always missing important reaction steps, and so give seriously erroneous predictions.

One might ask: how can one make sensible predictions when no data on one or more of the important reaction steps can be found in the literature? The answer is that chemists have abstracted a tremendous number of generalizations from the limited number of experimental (and, recently, theoretical) studies that have been performed on particular reactions. So when it is said that a particular reaction is “well-understood”, it is not meant that someone has actually measured that reaction’s rate to high precision under every possible reaction condition; instead it means that enough measurements and/or calculations have been done on that reaction or perhaps on some similar reaction so that one can reasonably generalize, and based on that generalization confidently estimate any particular parameter that is needed. Using generalizations has a tremendous advantage compared to attempting to list every possible individual reaction: one generalization can be used to estimate the parameters of hundreds or thousands of individual reactions. In fact, most of the rate parameters in the existing large kinetic models are actually estimates from generalizations, not from individual-reaction experiments or quantum calculations.

Most automated model-construction software uses only the conventional List-of-Reactions data model. Since there is no established data model for generalized rate estimates, nor for documenting the rules that determine which species and reactions are considered important, this information is typically not treated as “data” at all – instead, much of it is hard-coded into the software.

In our view, the main problem with existing software is that it is not well-designed to deal with the true complexity of the real chemistry, and corresponding leve of detail needed to accurately represent it, nor was it designed to be easily modified to incorporate new or improved chemistry knowledge. Because the chemistry has many details, and is imperfectly known, it is imperative that the software be easily extensible, so that additional chemical detail can be added as desired (preferably by the wide community of chemists). In this way, the model predictions can be continuously improved, each time incorporating the latest Chemistry Knowledge. If the software is not extensible, it will soon be obsolete.

II. The New Data Model for Kinetic Simulation

Here we propose a new data model that accurately represents the chemistry knowledge that goes into constructing a large model: a relatively small number of data on individual reactions which have been extensively studied in the literature plus a relatively small number of generalized rate and thermo estimation procedures which are used to estimate most of the thousands of parameters in the kinetic differential equations. We have developed a convenient data format for storing the rate and thermochemistry estimation parameters, as well as a graphical user interface, which makes it much easier to improve the rate estimates, to add additional reaction types, and to handle complicated functional groups. As we have shown elsewhere [Song (2004)], one can devise transparent unambiguous procedures, with clearly documented assumptions, for converting these estimation rules into lists of reactions, and thence into differential equations and quantitative predictions as is done currently. The new database format, graphical user interface, and algorithm for constructing the List of Reactions from our new tree-structured rate estimation database allows us to achieve the design shown in Fig. 2.

II. A. Hierarchical Tree Structure for Functional-Group Parameters

The information needed to compute both the thermochemical and the rate parameters can typically be associated with functional groups. To use these estimation approaches during mechanism construction, we need a reliable, efficient, and unambiguous method for rapidly identifying which functional group values should be used for any given molecule. To maximize scientific progress, we must display the functional groups and the numerical values we associate with each group in such a way that can be easily used, understood, scrutinized, criticized, amended, and extended by other researchers.

We have accomplished these goals by storing all the functional group definitions and the corresponding group values in a hierarchical tree database developed by R. Sumathi which is read both by the reaction mechanism generator RMG and by a graphical user interface (GUI) developed by J. Robotham. The RMG program performs substructure searches on the molecules in the model, to identify which species contain the functional groups in the tree database. Chemists can see and modify the rate estimates and functional group definitions in the database using the GUI, and then see how their changes affect the kinetic model that is constructed by RMG. The tree structure of the database facilitates efficient search for the best-matching functional group in each molecule encountered while the computer is building the reaction mechanism.

Although this concept seems straight-forward, implementation is complex particularly for polycyclic compounds, reactions through cyclic transition states, and species with unusual bonding. For more technical details see Song (2004). So far we have compiled more than 700 group values, and constructed hierarchical trees for more than 30 different types of chemical reaction.

II. B. The Reaction Mechanism Generator (RMG)

The new data model for chemical kinetics will work best if there is a reproducible, deterministic, automated procedure for going from the fundamental inputs (a library of the small number of known rate constants, and the hierarchical tree of rate estimation parameters for all the other reactions) to the desired outputs (e.g. predicted yield/selectivity profiles). Ideally, the software that performs this automated procedure will not need to be altered if someone changes an initial condition, a rate estimate, or even the structure of the hierarchical tree – this would dramatically reduce both the programming burden on the kineticist and the likelihood of introducing a bug into the software. This design also has the practical advantage that exactly the same software could be used by many kineticists for many different problems, making it much easier to reproduce work by another, while also reducing the difficulty of maintaining the software.

We have developed a software package, RMG, for exactly this purpose. The software is designed to take the hierarchical database trees as its main input, and to return a kinetic model as its main output. This kinetic model can be in the conventional List-of-Reactions format suitable for use in CHEMKIN or other integrators, or if desired the RMG program package can perform the integration itself and return product yield / selectivity profiles.

RMG, which was implemented by Jing Song, uses a modified version of the rate-based model-construction algorithm described in Susnow et al.(1997), but it treats non-thermalized chemically-activated reaction paths on an equal basis with ordinary thermal reactions, as done by Matheu et al. (2003). Byproducts and activated reaction intermediates are considered important species if their estimated instantaneous formation rates ever exceed R(t), where R(t) is the norm of the major species flux vector. The concentrations needed to compute the rates R(t) are estimated using the best mechanism available so far. The concentration estimates can be substantially in error early in the process when the mechanism may be missing important reactions. However, this procedure has been demonstrated to settle down and become quite stable and robust after a modest number of iterations even for rather complex systems [Matheu et al. (2003)]. At convergence, all of the neglected byproducts are formed at rates significantly lower than the rates of the important reactions in the system.

Unlike all previous model-generation software, RMG correctly handles pressure and temperature variations; it does this by using k(T,P) computed for the chemically-activated reactions at discrete (T,P) to determine coefficients in a Chebyshev form [Venkatesh et al. (1997)] suitable for use in the differential equation solver.

At the heart of the RMG algorithm is the process of identifying all the possible reactions of some species of interest X with itself and all the species {A,B,C…} included in the model so far, and estimating each reaction's rate. Unlike most previous mechanism generators, the RMG software handles this in a very general way: all the specifics about functional groups and reaction types, as well as all the associated parameters, are external to the program, contained in the tree databases. So nothing in the program has to change when the chemistry information is updated.

III. Results: A model for butane oxidation

The new RMG software was used to construct a model for the isothermal, isobaric oxidation of a stoichiometric mixture of butane and oxygen at T=715 K at the experimental conditions studied by Wilk *et al.*(1995). RMG constructed a model including more than 800 reactions. The predictions of this model are compared with the experimental data in Fig. 3.

The model and data do not match precisely (e.g. notice the discrepancy in the CH_4 yields) but it appears all the predicted yields are within about a factor of 4 of the experimental values, and most are much more accurate. The timescale of this complex cool-flame ignition process is also accurately predicted. One of course would not expect perfect agreement, since the experimental rate constants are seldom known to better than a factor of 2, and the uncertainties in the more than 700 group-additivity rate estimates required to build this model are considerably larger. As is often the case the predictions of the system kinetics are rather accurate despite all the uncertainties, since the kinetic system is only sensitive to the thermochemistry (well known for most of these compounds) and a handful of the rate constants.

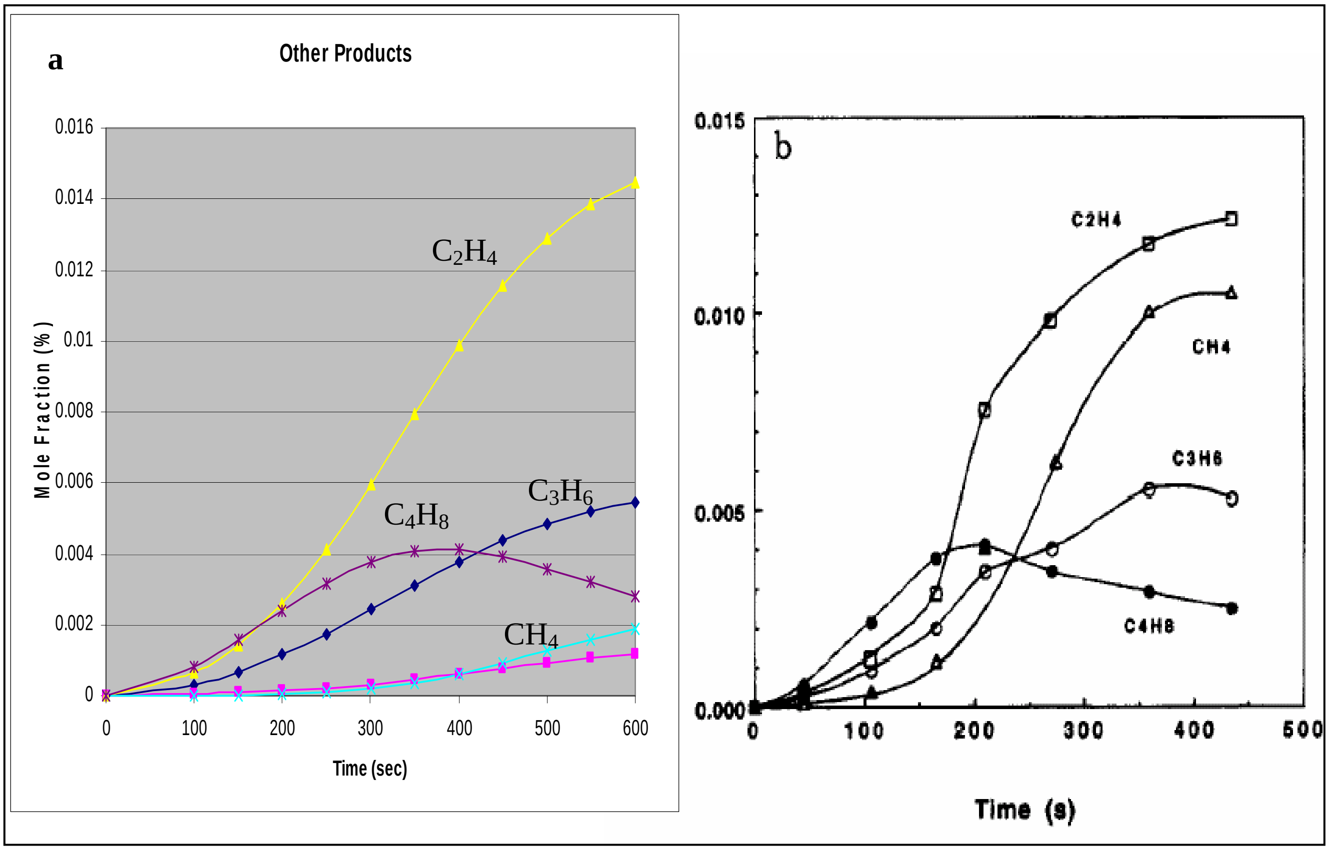

Fig. 3. (a) RMG predictions vs. (b) Wilk et al.'s experimental data on stoichiometric butane oxidation at 715 K.

IV. Summary and Conclusions

A new approach to constructing chemical kinetic simulations is presented, which is based on rate-estimation data upstream of the usual List-of-Reactions input data used in CHEMKIN and similar software. The new approach, Fig. 2, is greatly facilitated by use of new data formats, graphical user interfaces, and automated mechanism-generation software. A hierarchical tree data structure provides an easily extensible format well-suited to rate and property estimation based on functional groups. A graphical user interface and the RMG software make it practical to use and improve these databases. The new approach has been used to construct a kinetic model for butane oxidation in the cool-flame regime.

This paper argues that the fundamental data important for making chemical kinetic predictions is not just individual reaction rates, but also includes rate-estimation parameters derived as generalizations from the individual-reaction data. Hence, authors reporting a new rate constant value should be cognizant not only of prior studies of that particular rate constant, but also of rate constants for similar reactions, since these all taken together will be generalized to provide rate estimates.

The field of kinetics will advance most rapidly if researchers directly compare rate estimation procedures, in order to arrive at a consensus “best estimate”, and to identify discrepancies and holes in the data. The concept is similar to how some experimental rate constants have been evaluated. It is expected that the data structure presented here will facilitate these comparisons and improvements, and that the RMG software and other software using this new data model will provide an efficient and extensible way to put the rate estimates to practical use.

Bibliography

Benson, S.W. (1976). *Thermochemical Kinetics 2nd edition*, John Wiley & Sons, New York.

Kee, R.J.; Rupley, F.M.; Miller, J.A. (1989). “CHEMKIN-II: A FORTRAN chemical kinetics package for the analysis of gas-phase chemical kinetics.” Sandia National Laboratories Report, SAND89-8003. The modern versions of CHEMKIN are distributed commercially by ReactionDesign, Inc., http://www.reactiondesign.com

Matheu, D.M.; Dean, A.M.; Grenda, J.M.; Green, W.H. (2003). “Mechanism Generation with Integrated Pressure Dependence: A New Model for Methane Pyrolysis.” *J. Phys. Chem. A* **107**, 8552-8565.

Song, Jing. (2004). "Building Robust Chemical Reaction Mechanisms: Next Generatior of Automatic Model Construction Software”, Ph.D. dissertation, Massachusetts Institute of Technology.

Susnow, R.G.; Dean, A.M.; Green, W.H.; Peczak, P.; Broadbelt, L.J. (1997). “Rate-Based Construction of Kinetic Models for Complex Systems” *J. Phys. Chem. A* **101**, 3731-3740.

Tomlin A.S.; Turanyi, T; Pilling, M.J. (1997). “Mathematical Tools for the Construction, Investigation, and Reduction of Combustion Mechanisms.” in *Low Temperature Combustion and Autoignition* ed. by Pilling, M.J. & Hancock, G. *Comprehensive Chemical Kinetics* **35**:293*.* Elsevier, Amsterdam.

Venkatesh, P.K.; Chang, A.Y.; Dean, A.M.; Cohen, M.H.; Carr, R.W. (1997). ”Parameterization of pressure- and temperature-dependent kinetics in multiple well reactions”. *AIChE J.* **43**, 1331-1340.

Wilk, R.D.; Cohen, R.S.; Cernansky, N.P. (1995). “Oxidation of n-Butane: Transition in the Mechanism across the Region of Negative Temperature Coefficient.” *Ind. Eng. Chem. Res.* **34**, 2285-2297.