Mechanism Reduction during Computer Generation of Compact Reaction Models

David J. Klinke II and Linda J. Broadbelt
Dept. of Chemical Engineering, Northwestern University, Evanston, IL 60208

One of the challenges in building reaction mechanisms using algorithms for automated model construction is to describe the essential chemistry and enable prediction of experimental data over wide ranges of reaction conditions while maintaining a manageable model size. Two complementary methodologies for building compact reaction mechanisms were developed and combined with existing algorithms based on graph theory and bond–electron matrix operations. Each strategy was developed using pentadecylbenzene pyrolysis as an illustrative example. The first approach used a radical lumping strategy to group radicals according to their reactivity. The mechanism was reduced from 719 to 215 species and successfully predicted the experimentally observed initial reactivity. However, implementation of the radical lumping criteria alone was insufficient to allow for secondary reactions to higher-rank products. Therefore, on-the-fly sensitivity analysis was incorporated to identify the important and necessary species as the mechanism was generated to guide the mechanism building process. The generic algorithms developed can be applied to generate compact reaction mechanisms for a wide array of higher molecular-weight reactants.

Introduction

Economic needs and environmental legislation demand more explicit accounting of the composition of product streams in the chemical process and refining industries. As a result, the need for detailed kinetic models that facilitate this accounting has increased. Kinetic models provide a basis for reactor design and process improvement by allowing for fast and flexible exploration of diverse processing scenarios. They also provide a starting point for drawing inferences about the fundamental underlying, controlling chemistry.

For complex reacting systems, such as hydrocarbon feedstocks and oxidation chemistry, which can consist of $O(10^5)$ species, assembling knowledge of the chemistry into a kinetic model is formidable. Human resources limit the size of the model that can be realistically constructed, and hardware limitations restrict the size of the model that can be easily solved. These obstacles have motivated considerable interest in computer generation of reaction mechanisms that capture the essential features of the underlying chemistry. With such a tool, the modeler can focus on the information the model provides, rather than on the tedious manual construction of the model itself.

Although the algorithms that have been developed for the creation of molecular reaction models by computer and their implementations vary (Blurock, 1995; Broadbelt et al., 1994; Chevalier et al., 1990; Clymans and Froment, 1984; Di Maio and Lignola, 1992; Fontain and Reitsam, 1991; Hillewaert et al., 1988; Prickett and Mavrovouniotis, 1996a,b; Quann and Jaffe, 1992), many common features are found. The user provides a representation of the reactants' structure and information about the chemical reactions of interest. The structure of the reactants is transformed according to the chemical reactions to form reaction intermediates and products, and the connectivity and uniqueness of the products are determined. Finally, the relationships among the species are compiled into a reaction mechanism and the controlling differential equations that describe it.

The approach used in this work relies on earlier work by Broadbelt et al. (1994, 1995) in which the algorithms for reaction implementation, species' identification, and rate parameter estimation were developed. The implementation of reactions exploits graph theory and bond–electron matrix representations of molecules. The bond–electron matrix representation is a mathematically tractable representation of chemical species, where the *ij* entries of the matrix denote the bond

---

Correspondence concerning this article should be addressed to L. J. Broadbelt.

order between atoms *i* and *j*, and the *ii* entries denote the number of nonbonded electrons surrounding atom *i*. Chemical reaction is implemented through a simple matrix addition operation (Ugi et al., 1979), with a unique reaction matrix determined for each reaction type. The elements of the matrix are those values that, when added to the reactant(s) bond–electron matrix, effect the appropriate changes in the bond order and number of nonbonded electrons. Through repetitive application of the set of reaction matrices representing the chemistry of interest to the reactants and their progeny, the reaction mechanism is revealed. The reactions are carried out in a logical fashion to ensure that all species undergo their allowable reactions. When no new species that can undergo the specified reaction types are formed, mechanism generation halts.

The reaction mechanism has no value, however, without the values of the rate constants. Therefore, as a particular reaction is generated, the value of its rate constant is determined using a linear free-energy relationship (LFER). LFERs, in which the logarithm of the rate constant is related linearly to an appropriate reactivity index, provide estimates of rate constants using a single slope and an intercept for each reaction type. For the pyrolysis chemistry of interest in this work, heat of reaction is an appropriate reactivity index (LaMarca, 1992; Semenov, 1959). Thus, on-the-fly specification of rate constants requires only that the individual heats of reaction are calculated as the reaction is revealed.

The heats of formation of the species involved in a reaction are determined using a hierarchical approach. Preferentially, heats of formation are obtained through an interface to the National Institute of Standards Structure and Properties database (Stein et al., 1994; Susnow et al., 1996) where experimental values for a wide variety of species are tabulated, and a group additivity (Benson, 1976) approach is available. If the database fails to return a value, the heat of formation is calculated using MOPAC, a semiempirical computational chemistry package (Stewart, 1990).

These underlying algorithms have been implemented by Broadbelt et al. (1994, 1995, 1996) to generate manageable mechanisms for small molecular-weight reactants. However, one of the major challenges to successfully applying these generic tools to a wider array of chemistries and higher molecular-weight and multicomponent reactants has been unbounded mechanism growth in which the mechanism is too large to be easily analyzed or even solved. A variety of approaches has been developed to prevent this explosion in the number of species and reactions during computer generation of reaction mechanisms. The earliest implementation applied a heavy atom count, which placed a bound on the number of nonhydrogen atoms that a species capable of undergoing reaction could contain (Broadbelt et al., 1995). Application to higher molecular-weight reactants, however, revealed that the resulting species pool was skewed toward low molecular-weight compounds, as no lower bound prevented the formation of a large number of chemically insignificant, small species.

Implementation of a rank-based termination criterion resulted in more chemically meaningful reaction mechanisms (Broadbelt et al., 1995). The rank of a species is a measure of the order in which it appears as a product in the mechanism. The algorithm implemented formal mathematical rules to track a species' rank, and generated successive shells of molecular species of the same rank emanating from the set of reactants, as illustrated in Figure 1. However, application of these general rules for defining product rank still resulted in mechanisms that were not convergent. Radical addition reactions, which continually form radicals of higher molecular weight without any increment in rank, resulted in unbounded mechanism growth. Therefore, the rank-termination criterion was combined with a heavy atom count applied to radicals only and successfully prevented infinite mechanism generation. However, for chemistries involving important species of higher molecular weight (i.e., $> C_4$), the combined criteria were inadequate.

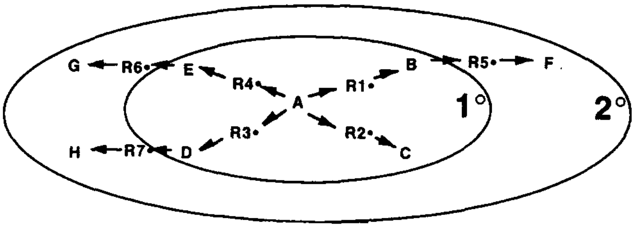

Figure 1. Rank-based termination criterion generates successive product shells with equal ranks.

Generation of a reaction mechanism for pyrolysis of a higher molecular-weight reactant, pentadecylbenzene, a model compound of heavy oils, provided one example of the inadequacy of these rules for mechanism termination. Initially, a mechanism was generated for pentadecylbenzene pyrolysis to form only primary products, and a carbon count of 25 was applied to radicals. This ensured that, minimally, all reactant-derived radicals were allowed to react. Even with these restrictive termination criteria, a mechanism comprising 719 components and over 6,000 reactions was generated and included a large number of molecular species that were not observed experimentally (yield $< 10^{-4}$).

It is apparent from this straightforward application that the value of the algorithms for computer generation of reaction mechanisms is limited by the lack of an appropriate scheme for termination such that compact, manageable reaction mechanisms that still maintain the essential chemistry are built. This deficiency motivated the present work, in which quantitative rate and concentration information is used during generation to determine the importance of a particular species, and therefore its ability to participate in subsequent reactions. By maintaining the inherent rank-based generation scheme outlined in Figure 1, the importance of a given species is evaluated through its interaction with other species of the same rank and lower. This contrasts with the recent work of Susnow et al. (1996), which also exploits quantitative values of rates and species' concentrations during mechanism generation. However, the mechanism is constructed by forming a small, initial pool of species, allowing only one species to react during a given iteration, and solving the mechanism after each iteration to evaluate the relative importance of the current set of species before the next iteration.

In this work, two approaches, each of which incorporates strategies similar to those applied to reduction of mechanisms in which all the species of interest are known *a priori*, were developed. However, these strategies are applied in an iterative fashion and guide the actual mechanism building process. The first approach utilizes an embedded lumping algorithm in which the species are examined for reactivity characteristics and are lumped into distinct reactivity classes as they are formed. This is analogous to approaches, for example, in which radicals are lumped according to their type (i.e., primary, secondary, tertiary) in long-chain pyrolysis modeling (Clymans and Froment, 1984). The second approach performs "on-the-fly" sensitivity analysis to direct the mechanism building process. Normalized Jacobian elements with respect to species' concentrations are evaluated through model solution for an initial mechanism, and the magnitude of the elements determines the next set of species allowed to react. In this way, the mechanism is iteratively solved and generated until a satisfactory description of the reaction chemistry is obtained. Each of these approaches is described in detail in the following sections using pentadecylbenzene pyrolysis as an illustrative example.

Embedded Lumping Strategy

The embedded lumping strategy was developed using pentadecylbenzene (PDB) pyrolysis as a representative case study. The structure of PDB suggests a large number of potential reaction pathways, but experimental data reveal that it reacts to form a small number of products during initial pyrolysis. The experiments and modeling carried out by Savage and Klein (1987, 1989) provided a target for mechanism building and were used to guide the selection of reactivity lumps. Savage and Klein (1987) studied the pyrolysis of PDB in constant-volume reactors in an inert argon atmosphere at temperatures ranging from 375 to 450°C, batch holding times ranging from 10 to 180 min, and initial concentrations ranging from 0.0043 to 2.3 M. Their experimental and modeling analyses were most comprehensive for the experimental data obtained at 400°C, and therefore, the development of the mechanism-generation algorithms used this data set as a guide.

The key experimental observation reported by Savage and Klein was that at low conversions, the products obtained from PDB pyrolysis could be lumped into two major product pairs, toluene and 1-tetradecene; and styrene and tridecane. A third lump, consisting of ethylbenzene and other n-alkanes, α-olefins, phenylalkanes, and phenylolefins, was minor. The presence of the two major product classes was accounted for by proposing the simple thermolysis pathways depicted in Figure 2. The formation of the product pairs toluene and 1-tetradecene, and styrene and tridecane, both in equimolar amounts, was consistent with initial selectivity data. To account for the disappearance of these major reaction products, styrene, 1-tetradecene, and n-tridecane were also allowed to undergo secondary decomposition reactions.

The ability of this compact set of reaction pathways to capture the observed experimental behavior simplified subsequent mechanistic analysis. Using classic Rice-Herzfeld pyrolysis chemistry comprising bond fission, propagation by hydrogen abstraction, β-scission, and radical recombination reactions, a mechanistic model was manually constructed. The key step in their manual construction of the model was to lump the radicals obtained by abstraction of hydrogen from PDB into three classes. The lumping of the PDB-derived radicals, denoted as the α-, β-, and γ-PDB-radicals, and depicted in Figure 3, was motivated by their distinct reactivities. The α-radical was unique because it was the most easily formed by hydrogen abstraction. Stabilization of the radical by the benzene ring resulted in a carbon-hydrogen bond strength of only 85 kcal/mol compared to 95 kcal/mol for all other alkyl carbon-hydrogen bonds. In contrast, the γ-radical was the most reactive radical, since upon β-scission, it forms a resonance stabilized benzyl radical through a bond cleavage of the β-PDB bond, requiring only 69 kcal/mol (Benson, 1976). At all other positions along the alkyl chain, the reactivity of the radical was essentially equivalent, so all other positions were lumped into one class identified as β-radicals. These radicals are still important, however, since they form α- and γ-radicals by chain transfer.

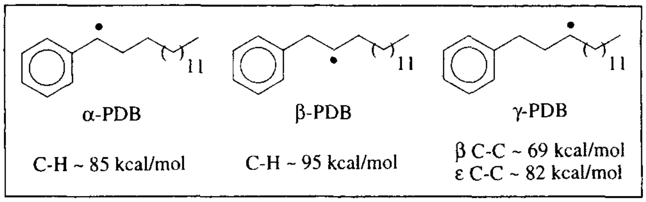

Figure 3. Three classes of pentadecylbenzene radicals lumped according to reactivity.

These important classes of PDB-radicals provided a starting point for strategic automatic mechanism generation. The mechanism generator was taught to identify the relevant structural attributes of the α-, β- and γ-PDB-radicals. To facilitate the classification, the aromatic ring, identified and labeled by the algorithm for finding cycles (Broadbelt et al., 1996), was used as a reference. For example, a PDB-radical with the radical center bonded to a carbon atom belonging to a cycle is uniquely the α-radical. Similar characteristics were identified to classify the γ-radical uniquely. If the radical center for a PDB-radical was not in the α- or the γ-position to the aromatic ring, it was classified as a β-radical. Although the radical assumes the identity of a β-radical, the reaction that forms it is added individually to the growing list of reactions to ensure that the proper reaction path degeneracy is maintained. This classification strategy overrides the isomorphism algorithm (Broadbelt et al., 1996; Corneil and Gotlieb, 1970), which would be able to distinguish among all 15 PDB-derived radicals. Instead, only three distinct radicals must be accounted for in subsequent reactions.

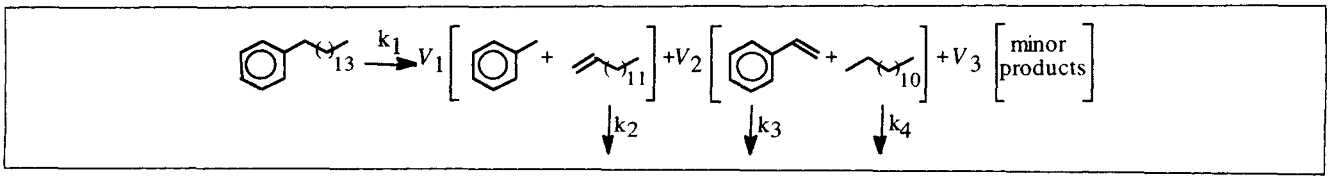

Figure 2. Pentadecylbenzene thermolysis pathway proposed by Savage and Klein (1987).

Table 1. Characteristics of Four Pentadecylbenzene Pyrolysis Mechanisms

Model	No. of Species	No. of Reactions	Characteristics
SK	51	121	Rice–Herzfeld
SK-MR	51	245	All reactions considered reversible
Full	719	6,030	Explicitly accounting for all products
NG	215	3,650	Employs 3 PDB radical classes

Application of the reactivity-based lumping strategy significantly reduced the size of the mechanism generated compared to the description of 719 species obtained without the embedded lumping algorithm. A more compact mechanism of only 215 species and 3,650 reactions resulted when only primary products were generated, and a carbon count of 25 applied to radicals was used.

It is apparent from the work of Savage and Klein that the radical lumping scheme combined with a simple Rice–Herzfold pyrolysis mechanism was capable of capturing the major features of the initial reactivity of PDB. Their manually constructed model contained 51 species and 121 reactions and was solved using representative parameters for activation energies and preexponential factors. It is clear, however, from a comparison of the number of species and reactions that the computer-generated mechanism incorporating the radical lumping strategy and typical free-radical reaction types was not identical to that of Savage and Klein. The use of LFERs and an individual heat of reaction (i.e., individual activation energy) for each reaction distinguished the models further. The models were therefore directly compared through model solution to quantify the effect of the additional species and reactions and the use of LFERs to capture the kinetics.

Overall, three mechanisms were evaluated and compared. The relatively compact Rice–Herzfeld mechanism of Savage and Klein was manually reproduced and will be referred to as the SK model. In order to provide a more direct comparison with the computer-generated mechanism, which ensures that all reactions in the mechanism are reversible, the SK model was manually manipulated to include the reverse pair of each reaction. This model is designated as SK-MR. The third mechanism was the one generated automatically using the embedded-radical lumping scheme and is denoted as NG. The mechanism characteristics are summarized in Table 1, and although it was too large to be examined in detail because of hardware limitations, the characteristics of the mechanism generated without lumping are included for comparison.

The models were each solved and compared using reactant conversion and major product yields and selectivities as the key measures of model performance. The parameters used to solve each model are summarized in Table 2. The SK and SK-MR parameters are the values used by Savage and Klein, and were therefore used without subsequent refinement. The parameters used to solve the NG model comprised frequency factors and LFER parameters. The α values were fixed at physically reasonable values (Nigam et al., 1992) and constrained to ensure thermodynamic consistency between reverse pairs of reactions. The E_0 values were initially fixed to match the activation energies used in the SK model for a representative reaction in each reaction family. The frequency factors were also initially set to the values used in the SK model. Subsequent refinement of a subset of these parameters (shown in boldface in Table 2) through global optimization (Stark, 1993) using experimental data up to 30 min of reaction time resulted in the final set of NG parameters summarized in Table 2. These optimized parameters deviated overall only slightly from the initial guesses suggested by the SK parameter set.

The predicted initial selectivity of toluene to styrene, two important major products, as a function of initial PDB concentration for the SK, SK-MR, and NG mechanisms is compared to the experimental data of Savage and Klein (1987) in Figure 4. The SK and NG mechanisms predict the dependence of the initial selectivity on the initial PDB concentration quite well. However, the SK-MR model dramatically overpredicts the selectivity at higher concentrations. Since no adjustment of the SK parameters was used in the solution of SK-MR, it is not possible to judge its overall predictive capability. However, it is apparent that including the reverse reactions using physically reasonable parameters has a significant impact on one key measure of model performance. More critical to this work, it is clear that the more comprehensive computer-generated mechanism incorporating LFERs and the embedded-radical lumping strategy reproduces the initial selectivity of toluene to styrene well.

Calculation of the overall pseudo-first-order rate constants for PDB disappearance provided a second comparison among

Table 2. Arrhenius and Linear Free-Energy Relationship Parameter Values Used to Solve the Three Different Models Examined

Model		Initiation	*H*-Abstraction	Recombination	*β*-Scission	Addition
SK	A ($\log_{10}$)	16	8.5	8.5	15	—
	E (kcal/mol)	68	14/18*	0	28/38**	—
SK-MR	A ($\log_{10}$)	16	8.5	8.5	15	9
	E_0 (kcal/mol)	68	14/18*	0	28/38**	15
	α	—	—	—	—	0.5
NG	A ($\log_{10}$)	**15.87**	**8.49**	**8.97**	**14.58**	**9.87**
	E_0 (kcal/mol)	0	**14.85**	0	**18.6**	**18.6**
	α	1	0.5	0	0.5	0.5

*14 kcal mol^{-1} for α-carbon, 18 kcal mol^{-1} for non-α-carbon.
**28 kcal mol^{-1} for γ-radical, 38 kcal mol^{-1} for non-γ-radical.

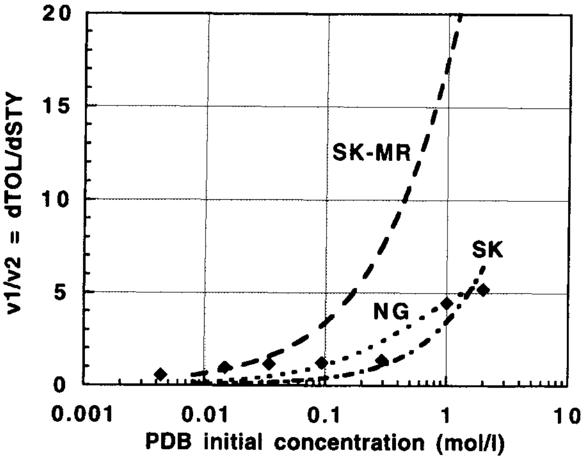

Figure 4. Initial selectivity of toluene to styrene as a function of initial PDB concentration for SK, SK-MR, and NG mechanisms.

◆, experimental selectivity (Savage and Klein, 1987).

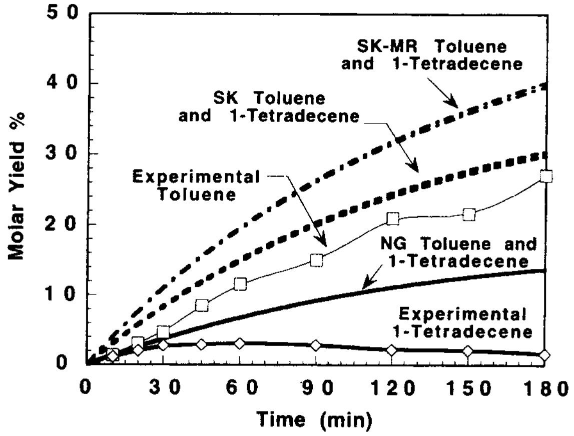

Figure 6. NG model predicts the experimental yields of toluene and 1-tetradecene at 400°C at low reaction times well.

the three mechanisms. The SK, SK-MR, and NG model predictions are compared with the experimental values as a function of initial PDB concentration in Figure 5. Although the predictions of the three mechanisms differ slightly, it appears that including the reverse reactions is important to capture the overall disappearance kinetics. The NG and SK-MR models predict the experimentally observed curvature as a function of initial concentration slightly better than the SK model. Overall, the more comprehensive NG model provides the best prediction over the full range of concentrations.

Further evaluation of the performance of the computer-generated mechanism for PDB pyrolysis was carried out by comparing the predicted and experimental temporal yields of major products. Comparison of the temporal variations of toluene and 1-tetradecene predicted by the NG model to the experimental data is made in Figure 6. The NG model clearly predicts the initial rates of formation of both toluene and 1-tetradecene at reaction times less than 30 min quite well. At longer times, deviation of the model results from the experimental data is significant. After approximately 30 min, secondary pathways for both products are important, as the yield of 1-tetradecene decreases, and the yield of toluene continues to increase. The disagreement at longer times is not surprising, however, since these secondary reaction pathways are not included in the NG mechanism. As shown in Figure 6, the SK and SK-MR models also fail to capture this behavior, since neither includes secondary reactions to products of higher rank. In order to address these deficiencies, the logical next step was to incorporate secondary reactions into the existing SK, SK-MR, and NG models. Since the SK and SK-MR models were manually constructed, however, it would be extremely tedious to extend these primary mechanisms to include all secondary reactions due to the combinatorial nature of the chemistry. Therefore, only the NG mechanism was expanded, facilitated by the algorithms for automatic model construction.

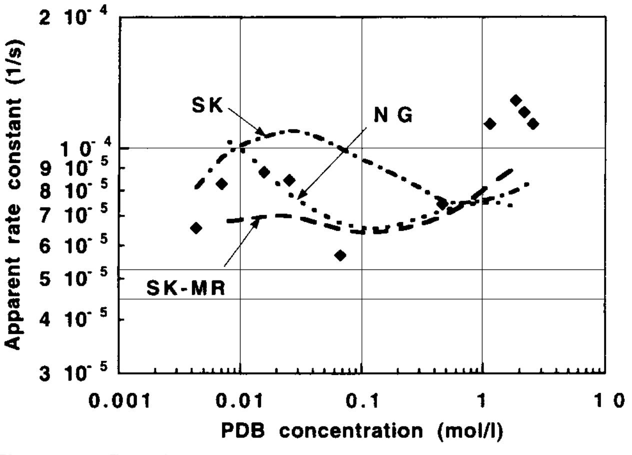

Figure 5. Predicted pseudo-first-order rate constants as a function of initial pentadecylbenzene concentration for SK, SK-MR, and NG mechanisms.

◆, experimental selectivity (Savage and Klein, 1987).

Using the embedded-radical lumping strategy, generation of a mechanism for PDB pyrolysis forming up to secondary products was attempted. However, hardware limitations encountered due to the memory required to store the species information, particularly large products formed by radical recombination, prevented successful generation of the model. At the point where computational resources were exhausted, over 10^5 reactions had been revealed, and 2 GB of memory were reserved. Note that these hardware limitations would be encountered at an earlier stage in mechanism generation if formation of secondary products was allowed without the embedded radical lumping strategy.

It is apparent at this point that the embedded radical lumping strategy successfully reduced the primary product mechanism to a manageable size and still allowed the essential features of the initial reactivity during PDB pyrolysis to be captured as compared to experimental data. However, additional criteria for model reduction during generation are required to allow for the formation of secondary and higher-order reaction products. It is desirable to incorporate termination criteria that allow for the inclusion of kinetically important reactions without sacrificing the chemical detail already present. Therefore, "on-the-fly" sensitivity analysis was incorporated to direct the mechanism-building process and is described in detail below.

Generation of Reaction Mechanisms Involving Important and Necessary Species

Important and necessary species

On-the-fly application of sensitivity analysis to direct the mechanism-building process draws upon the work of Turányi (1990b), who developed strategies for mechanism reduction when all of the species of interest are known *a priori*. Turányi (1990b) classified species into three different categories: important, necessary, and redundant. This terminology is adopted to describe the approach used herein. Important species are those for which the accurate reproduction of the concentration profiles follows directly from the aim of the investigation. The decision as to which species and features are considered important depends on the modeling objective. In this pentadecylbenzene pyrolysis example, it is critical that the disappearance of PDB and the temporal variations of the two major product pairs are reproduced. Necessary species are those for which realistic concentrations are required in order to calculate accurate concentration profiles for the important species or to reproduce important features. Redundant species are those that can be eliminated from the model without jeopardizing the aims of the modeling. It is particularly important in the automatic generation of networks to identify redundant species at an early stage to minimize extraneous mechanism growth.

The approach used by Turányi (1990b) to identify redundant species was incorporated directly into algorithms for automatic network generation. Specifically, a species may be considered redundant if its concentration change, dc_i, has no significant effect on the net rate of production of important species. The elements of the normalized Jacobian, $(\partial \ln f_n(t))/(\partial \ln c_i(t))$, are calculated as a measure of such an effect, where $f_n(t)$ is the net rate of change of species n. The overall influence of the change of the concentration of species i on the rate of production of an N-membered group of important species is quantified by the sum of squares of normalized Jacobian elements (Eq. 1) and captured in a single parameter, $B_i(t)$:

$$B_i(t) = \sum_{n=1}^{N} [\partial \ln f_n(t)/\partial \ln c_i(t)]^2. \qquad (1)$$

Large values of $B_i(t)$ indicate a significant effect of species i on the set of important species. Alternatively, small values of $B_i(t)$ indicate that a species has little or no effect on the important species, and reactions in which the redundant species is a reactant may be safely removed.

The overall determination of the subset of important and necessary (IN) species within the mechanism generation algorithm is an iterative process as the mechanism is built. The essential idea is to generate a compact primary product mechanism and apply the criteria to determine those species that are IN based on an initial list of important species supplied to the mechanism generator. The $B_i(t)$ values are a measure of the *direct* effect of species i on the current list of IN species. However, there are necessary species that are linked to the important species through other necessary species (Turányi, 1990b). Therefore, the current IN species list is iteratively updated for a mechanism of a given rank until convergence is achieved.

After the IN species comprising the generated mechanism of a particular rank have been identified, the mechanism of that same rank is generated anew. However, only those species that were deemed IN are allowed to react, thereby eliminating any reactions in which redundant species were reactants. Removal of these reactions also eliminates the formation of other redundant species contained in the original mechanism, and therefore, the total number of species comprising the mechanism of a given rank is reduced. Next, the mechanism generation continues to form the set of products of the next higher rank, again allowing only the IN species to participate in reactions. The iterative process again continues until no new species are identified as important and necessary. Although it is impossible to ensure that a higher-rank species formed at a later time would not have caused a previous species identified as redundant to become important or necessary, the pyrolysis mechanisms of interest in this work are not significantly nonlinear. In instances where more feedback occurs, such as that observed in oxidation chemistry, the initial list of IN species can be expanded to prevent the early removal of a species that would be deemed important or necessary later.

The algorithm for determining the IN species during automatic mechanism generation is detailed as follows:

```
BEGIN
  Read in initial list of important species to IN array;
  Read in reactants to unreacted species list;
  Generate primary mechanism based on unreacted species list;
  FOR mechanism rank = 1 to final rank
    Read in initial concentrations;
    Initialize important species list to original list;
    Initialize differential algebraic equation solver;
    FOR index = 1 to number of output times
      Calculate y and dy/dt for all species;
      Set number IN species old equal to zero;
      Set number IN species equal to number in important species list;
      IF number IN species old = number IN species THEN
        number IN species old = number IN species;
        FOR species index = 1 to number of species
          Calculate $B_i(t)$ where $i$ = species index;
          IF $B_i(t) >$ threshold value THEN
            Add species to list of IN species;
            number IN species = number IN species + 1;
          END IF {$B_i(t)$}
        CONTINUE {species index}
      END IF {number IN species old}
      Add new IN species to IN array;
    CONTINUE {index}
    Delete all lists of molecules and reactions;
    Reinitialize unreacted species list;
    Generate mechanism of defined rank;
  CONTINUE {mechanism rank}
END Algorithm
```

The first step is to generate an initial mechanism using carbon-count and rank-termination criteria for the reactants in the initial input list. This mechanism is solved to obtain species concentrations and rates at a specified number of output times. The $B_i(t)$ values, as defined in Eq. 1, are evaluated for each species and compared to a threshold value that has been provided as input. If the resulting $B_i(t)$ value is greater than the threshold, species i is added to the set of IN species. This implies that for the next species, the summation of the Jacobian elements will be over one additional IN species. Furthermore, the determination of the set of IN species is an iterative approach, where control is passed to the next output time only when no new IN species are discovered after looping through the entire species list. As a result, a species that was not deemed IN in a previous iteration may become IN when the updated list is used, even when the output time is fixed.

The evaluation of the IN species is conducted at each specified time step. The final set of IN species for the mechanism examined is the union of the sets of IN species determined at each individual time step. Once this final set of IN species is determined, the initial mechanism is purged from the stored memory, and the initial reactant list is reinitialized. The allowable product rank is incremented, and the mechanism is generated again, but only allowing the IN species to react. This process facilitates the removal of those reactions in which a redundant species is a reactant, resulting in reduction of the mechanism at a particular rank with little added computational burden. The algorithm is then repeated at a sequence of product ranks until the final desired rank is achieved.

Impact of user-specified parameters

PDB pyrolysis was used as a case study to measure the sensitivity of the generated mechanism to the three main user-specified parameters: the set of the initial IN species, the threshold value, and the array of output times. Of the three variables, the set of initial IN species offers the least flexibility since the modeling goals are typically well defined. In the case of PDB pyrolysis, accurate prediction of the yields of the major products toluene, styrene, tridecane, 1-tetradecene, and ethyl benzene and the reactant PDB was desired. Therefore, the initial list of IN species comprised these six species for all mechanisms generated.

The threshold value had a profound impact on the determination of the final set of IN species. This effect was anticipated since in previous applications of this method to the reduction of manually constructed mechanisms, Tomlin et al. (1992) found that the $B_i(t)$ values were distributed, and a natural break point between the redundant and IN species was observed. For PDB pyrolysis, the distribution of $B_i(t)$ values for threshold values of 1.80 and 300 are shown in Figures 7a and 7b, respectively, in which the frequency of the $B_i(t)$ values in a particular range vs. the logarithm of the $B_i(t)$ value is plotted. The $B_i(t)$ values reported represent an average for a particular species over all reaction times examined. The distribution of Figure 7a is clearly bimodal and reveals that only those species in the upper portion of the distribution will be important and necessary, as indicated by the line of demarcation. However, the number of IN species cannot be explicitly tailored by simply examining the distribution and selecting the threshold value that divides the species into the two desired groups. In the application of the IN criteria to static mechanisms as carried out by Turányi (1990a,b) and in computer-generated reaction mechanism building, the determination is iterative, with the IN species pool over which the Jacobian elements are calculated developing dynamically. Therefore, the final number of IN species and the threshold value are not independent. This is indicated clearly by the histogram of Figure 7b, in which the threshold value was increased to 300. The number of IN species decreased to 28, a number less than the pool of 64 species expected from an examination of Figure 7a.

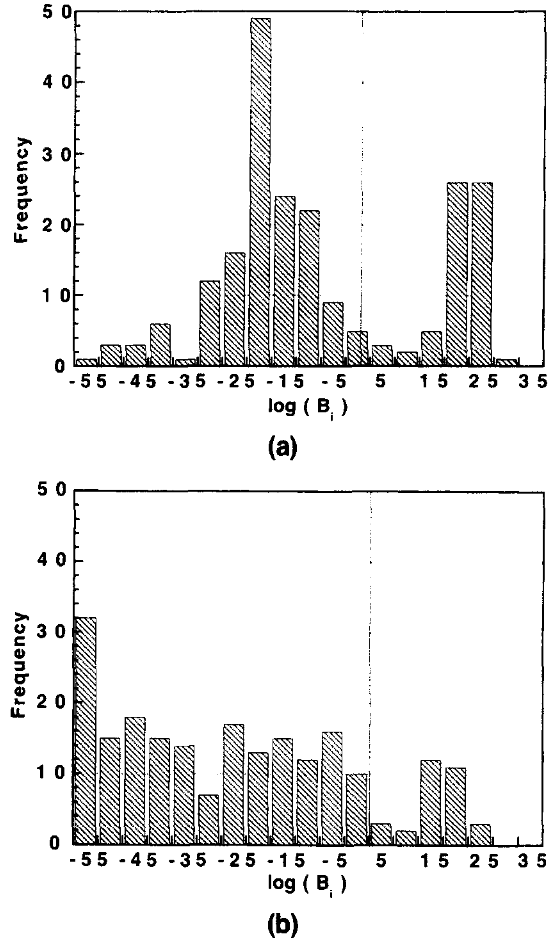

Figure 7. Distribution of B_i values for pentadecylbenzene pyrolysis for a threshold value of (a) 1.80; (b) 300.

To examine the dependence of the final number of IN species on the threshold value, the threshold value was systematically varied over many orders of magnitude. The number of IN species as a function of the logarithm of the threshold value is plotted in Figure 8. In general, as the threshold value is increased, there is a monotonic decrease in the number of IN species until a threshold of $O(10^{-4})$ is reached, and a discernible plateau is observed. As indicated by the relatively steep decline after 10^0 from a value of 64 to 6, there is a critical range of threshold values above which no additional species are included as IN species except those specified initially. At the other extreme, the threshold must be decreased to 1×10^{-50} in order to denote all 215 species involved in the mechanism as important and necessary.

The array of output times used to evaluate the IN species had an observable effect on the final set of IN species. A plot of the species number assigned during mechanism generation

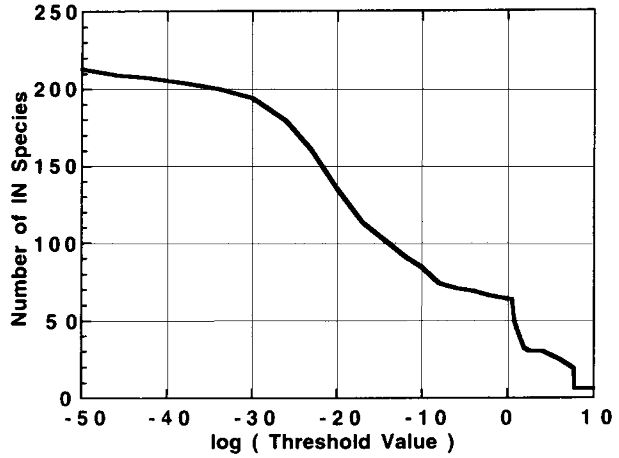

Figure 8. Number of IN species determined for pentadecylbenzene pyrolysis is a strong function of the user-specified threshold value.

as a function of time for primary PDB pyrolysis is shown in Figure 9, where a solid dark mark is used to denote that the particular species was identified as IN at the corresponding output time. The set of IN species evolved with time, with more species included as time, or reactant conversion, increased. Turányi (1990a) observed a similar dependence of the IN species set on reactant conversion. The majority of species with low identification numbers are important and necessary at low reactant times as well as at high. This is consistent with the mechanism-generation algorithm, which generates species that come directly from the reactant, and would therefore impact its initial disappearance, first. Other species described by the primary PDB mechanism are only IN at long reaction times. However, it was already shown that the primary mechanism with all species included provided an inadequate description of the chemistry at reaction times greater than 30 min. Therefore, the maximum output time, and therefore the maximum conversion achieved, is increased simultaneously as the product rank used to specify the current mechanism is increased.

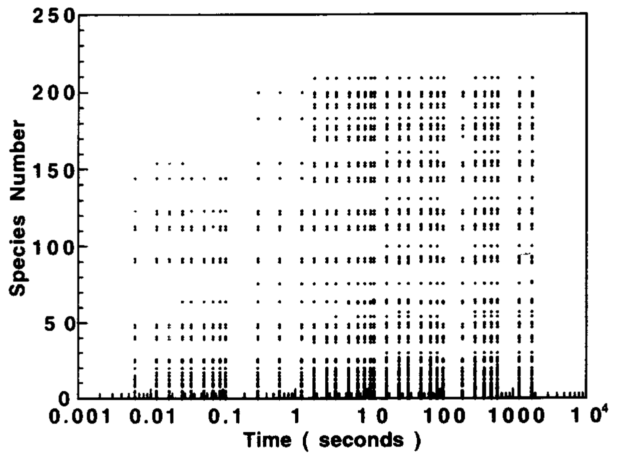

Figure 9. Identified IN species designated by a dark point at each specified time step.

Comparison of pentadecylbenzene pyrolysis mechanisms

The mechanism generation strategy incorporating the concept of important and necessary species both with and without the embedded radical lumping algorithm was applied to pentadecylbenzene pyrolysis. The initial IN species list was comprised of the five major reaction products and the reactant PDB, and the threshold was established just below the critical threshold value observed in Figure 8. The characteristics of the resulting primary mechanisms are summarized in the second and third columns of Table 3. For the mechanism without the radical lumping strategy (i.e., Full), 117 species were identified as IN. When the radical lumping strategy was also included, only 64 species were IN. Note that identification of the IN species for the primary mechanism had no impact on the overall mechanism characteristics. However, this was not surprising since all of the rank-zero radicals were identified among the IN species in both cases. Since this is exactly the set of species allowed to react in the primary mechanism in addition to PDB, no reactions involve redundant species as reactants, and therefore, none are removed.

The value of the IN species identification was clear when the mechanisms were expanded to allow for the formation of secondary products. For the Full mechanism, both with and without the IN criterion, and for the mechanism incorporating radical lumping but not the IN criterion, generation of a mechanism including higher-rank products failed. In some cases, available memory was exhausted and in others, a projection based on the number and types of species yet to react indicated a prohibitively large mechanism. In was only possible to generate a mechanism allowing for secondary reaction pathways when both the embedded-radical lumping strategy and the identification of IN species were incorporated. These results clearly indicate that a combined approach using structure and reactivity information and quantitative evaluation of reaction rates provides the most effective strategy for computer generation of chemically meaningful reaction mechanisms for higher molecular-weight reactants.

The next logical step was to compare the predictions of the model constructed (IN/RL), which allowed for the consumption of primary products to form secondary products, to the experimental data. The comparison primarily focused on the divergence of the yields of the product pair, toluene and 1-tetradecene, at reaction times greater than 30 min, since this was the dominant feature of the experimental data that the kinetic models allowing for the formation of only primary products fails to capture. The IN/RL model was solved using

Table 3. Incorporation of Both the Rules for Embedded Radical Lumping and the IN Criteria Resulted in the Only Manageable Mechanism that Allowed for the Formation of Secondary Products

	Primary Products		Secondary Products	
Mechanism	Species	Reactions	Species	Reactions
Full	719	6,030	2,600 +*	27,000 +*
IN Reduced Full	719 (117)**	6,030	1,060 +*	12,500 +*
Lumped	215	3,650	460 +*	26,000 +*
IN Reduced Lumped	215 (64)**	3,650	2,160	64,691

*Hardware limitations prohibited determination of exact values.
**Number of IN species.

Table 4. Comparison of Molar Ratios of Toluene to 1-Tetradecene Observed Experimentally to Prediction of IN/RL Model

Reaction Time (min)	Exp. Molar Ratio of Toluene to 1-Tetradecene	IN/RL Model Prediction of Molar Ratio of Toluene to 1-Tetradecene
10	1.25	1.70
20	1.5	2.12
30	1.67	2.49
45	2.89	3.05
60	3.71	3.61
90	5.28	4.73
120	9.42	5.89
180	10.33	7.07

the NG parameters of Table 2, with the frequency factor for radical addition increased to 3.4×10^{11} $M^{-1}\cdot s^{-1}$. The yields of toluene and 1-tetradecene predicted by the IN/RL model did diverge due to the presence of secondary reactions. The model results are compared to the experimental values in Table 4 in terms of the ratio of the moles of toluene to the moles of 1-tetradecene, which is an increasing function of reaction time. These results are clearly a significant improvement upon the NG, SK, and SK-MR models that did not predict any difference in the toluene and 1-tetradecene yields over the full range of reaction times examined. It was noted that the PDB conversions predicted by the IN/RL model were too low compared to the experimentally observed values. However, the prediction was improved when the threshold for model construction was decreased to allow more species to participate in primary and secondary reactions, thus consuming both PDB and 1-tetradecene. A second strategy for improving the overall agreement between the model and the experimental results that was not explored in this work is inclusion of additional reaction types during model construction. A reaction pathway such as intramolecular radical addition would provide an additional consumption pathway for 1-tetradecene-derived radicals.

Conclusions

The major aim of the present work was to develop methodologies for building compact reaction mechanisms within the bounds of current computational hardware limitations that still maintain the essential chemistry and allow prediction of experimental data over wide ranges of conversion and reaction conditions. Two complementary approaches were incorporated into existing software for computer generation of reaction mechanisms. The first approach used a radical lumping strategy to group radicals according to their reactivity. The mechanism generator was taught to identify the relevant structural attributes that unambiguously distinguished among radicals with distinct reactivity. Lumping species reduced the combinatorial mechanism growth observed for pyrolysis chemistry while the appropriate reaction path degeneracy was maintained. The radical lumping strategy reduced the mechanism size from 719 to 215 species and predicted successfully the experimentally observed initial reactivity. However, the implementation of the radical lumping criteria alone was not sufficient to allow for secondary reactions to higher-rank products. Therefore, on-the-fly sensitivity analysis was incorporated to identify the IN species as the mechanism was generated to guide the mechanism-building process. It was found necessary to combine both strategies to generate a mechanism including secondary products in order to maintain a reasonable mechanism size that incorporated the essential chemistry and to overcome current hardware limitations. The generic elements of the algorithms developed can be applied to generate compact reaction mechanisms for a wide array of higher molecular-weight reactants.

Acknowledgments

The authors are grateful to Exxon Research and Engineering Company for their financial support of this work.

Literature Cited

Benson, S. W., *Thermochemical Kinetics*, Wiley-Interscience, New York (1976).

Blurock, E. S., "Reaction: System for Modeling Chemical Reactions," *J. Chem. Inf. Comput. Sci.*, **35**, 607 (1995).

Broadbelt, L. J., S. M. Stark, and M. T. Klein, "Computer Generated Pyrolysis Modeling: On-the-Fly Generation of Species, Reactions, and Rates," *Ind. Eng. Chem. Res.*, **33**, 790 (1994).

Broadbelt, L. J., S. M. Stark, and M. T. Klein, "Termination of Computer-Generated Reaction Mechanisms: Species Rank-Based Convergence Criterion," *Ind. Eng. Chem. Res.*, **34**, 2566 (1995).

Broadbelt, L. J., S. M. Stark, and M. T. Klein, "Computer Generated Reaction Modelling: Decomposition and Encoding Algorithms for Determining Species Uniqueness," *Comput. Chem. Eng.*, **20**, 113 (1996).

Chevalier, C., J. Warnatz, and H. Melenk, "Automatic Generation of Reaction Mechanisms for Description of Oxidation of Higher Hydrocarbons," *Ber. Bunsen-Ges. Phys. Chem.*, **94**(11), 1362 (1990).

Clymans, P. J., and G. F. Froment, "Computer-generation of Reaction Paths and Rate Equations in the Thermal Cracking of Normal and Branched Paraffins," *Comput. Chem. Eng.*, **8**, 137 (1984).

Corneil, D. G., and C. C. Gotlieb, "An Efficient Algorithm for Graph Isomorphism," *J. Assoc. Comput. Mach.*, **17**, 51 (1970).

Di Maio, F. P., and P. G. Lignola, "KING, A Kinetic Network Generator," *Chem. Eng. Sci.*, **47**, 2713 (1992).

Fontain, E., and K. Reitsam, "The Generation of Reaction Networks with Rain. 1. The Reaction Generator," *J. Chem. Inf. Comput. Sci.*, **31**, 96 (1991).

Hillewaert, L. P., J. L. Dierickx, and G. F. Froment, "Computer Generation of Reaction Schemes and Rate Equations for Thermal Cracking," *AIChE J.*, **34**, 17 (1988).

LaMarca, C., "Kinetic Coupling in Multicomponent Pyrolysis Systems," PhD Diss., Univ. of Delaware, Newark (1992).

Nigam, A., C. LaMarca, D. Fake, and M. T. Klein, "Semiempirical Rate Laws for Rice-Herzfeld Pyrolysis of Mixtures: Capturing Chemistry with Reasonable Computational Burden," *Energy Fuels*, **6**, 845 (1992).

Prickett, S. E., and M. L. Mavrovouniotis, "Construction of Complex Reaction Systems: 1. Reaction Description Language," *Comput. Chem. Eng.*, in press (1997a).

Prickett, S. E., and M. L. Mavrovouniotis, "Construction of Complex Reaction Systems: 2. Molecule Manipulation and Reaction Application Algorithms," *Comput. Chem. Eng.*, in press (1997b).

Quann, R. J., and S. B. Jaffe, "Structure-oriented Lumping: Describing the Chemistry of Complex Hydrocarbon Mixtures," *Ind. Eng. Chem. Res.*, **31**, 2483 (1992).

Savage, P. E., and M. T. Klein, "Asphaltene Reaction Pathways. 2. Pyrolysis of n-Pentadecylbenzene," *Ind. Eng. Chem. Res.*, **26**, 488 (1987).

Savage, P. E., and M. T. Klein, "Kinetics of Coupled Reactions: Lumping Pentadecylbenzene Pyrolysis into Three Parallel Chains," *Chem. Eng. Sci.*, **44**, 985 (1989).

Semenov, N. N., *Some Problems of Chemical Kinetics and Reactivity*, Pergamon Press, New York (1959).

Stark, S. M., *MLSL: Multi Level Single Linkage Global Optimization Program Manual*, Newark, DE (1993).

Stein, S. E., S. G. Lias, J. F. Liebman, R. D. Levin, and S. A. Kafafi, *NIST Structures and Properties Users' Guide*, U.S. Dept. of Commerce, National Institute of Standards and Technology, Gaithersburg, MD (1994).

Stewart, J. J. P., "Semiempirical Molecular Orbital Methods," *Reviews in Computational Chemistry*, K. B. Lipkowitz and D. B. Boyd, eds., VCH Publishers, New York (1990).

Susnow, R. G., A. M. Dean, W. H. Green, P. Peczak, and L. J. Broadbelt, "Rate-based Construction of Kinetic Models for Complex Systems," *J. Phys. Chem. A.*, **101**(20), 3731 (1997).

Tomlin, A. S., M. J. Pilling, T. Turányi, J. H. Merkin, and J. Brindley, "Mechanism Reduction for the Oscillatory Oxidation of Hydrogen: Sensitivity and Quasi-Steady-State Analyses," *Combust. Flame*, **91**, 107 (1992).

Turányi, T., "Reduction of Large Reaction Mechanisms," *New J. Chem.*, **14**, 795 (1990a).

Turányi, T., "Sensitivity Analysis of Complex Kinetic Systems. Tools and Applications," *J. Math. Chem.*, **5**, 203 (1990b).

Ugi, I., J. Bauer, J. Brandt, J. Friedrich, J. Gasteiger, C. Jochum, and W. Schubert, "New Applications of Computers in Chemistry," *Angew. Chem. Int. Ed. Engl.*, **18**, 111 (1979).

Manuscript received Nov. 15, 1996, and revision received Mar. 17, 1997.