Application of Chemical Graph Theory for Automated Mechanism Generation

Artur Ratkiewicz*,[†,‡] and Thanh N. Truong*,[†]

Henry Eyring Center for Theoretical Chemistry, Department of Chemistry, University of Utah, 315 South
1400 East, Room 2020 Salt Lake City, Utah 84112, and Chemistry Institute, University at Bialystok, Al.
Pilsudskiego 11/4 15-443 Bialystok, Poland

We present an application of the chemical graph theory approach for generating elementary reactions of complex systems. Molecular species are naturally represented by graphs, which are identified by their vertices and edges where vertices are atom types and edges are bonds. The mechanism is generated using a set of reaction patterns (sub-graphs). These subgraphs are the internal representations for a given class of reaction thus allowing for the possibility of eliminating unimportant product species a priori. Furthermore, each molecule is canonically represented by a set of topological indices (Connectivity Index, Balaban Index, Schulz TI Index, WID Index, etc.) and thus eliminates the probability for regenerating the same species twice. Theoretical background and test cases on combustion of hydrocarbons are presented.

1. INTRODUCTION

A critical bottleneck in bridging the gap between space and time scales from individual reactions (length scale of nanometers and time scale of femto-seconds) to reacting flow (length scale of meters and time scale of minutes to hours) is a proper description of the mechanisms of complex systems. One of the most important complex processes in industry is combustion of hydrocarbons. The ability to perform detailed simulations of such processes would have significant impacts on many industries as well as on our environment. To achieve this goal one must be able to develop detailed kinetic models of complex combustion reactions. In practical combustion systems, kinetics models are in the order of thousands of elementary reactions or more and a large number of reactive intermediates. For example, there are 3662 reactions involving 470 species considered in simulations of *n*-hexane combustion by Come and co-workers[1] and 479 206 reactions and 19 052 species in simulations of tetradecane combustion performed by Broadbelt and co-workers.[2] Such kinetic models require a tool capable of automatically generating a mechanism based on a known set of reaction types. According to an excellent review of Tomlin et al.[3], a mechanism generation program should have the following features:

(1) Generated chemical species should be stored in a form that can be easily manipulated and its notation must be unique and nonambiguous (canonical).

(2) The algorithm must identify given species in the same way each time it is created (if species or reaction is formed more than once, an appropriate multiplicity should be increased).

(3) It should be able to filter out those reactions which are obviously unimportant (that is, very slow and/or with high activation energy, that could be eliminated a priori).

Several algorithms for generating a mechanism have been developed. They differ in the external and internal representations of species and reactions and in the algorithm for determining the canonicity of the generated species. Chinnick et al.[4] were among the first to develop mechanism generators based on logical programming. Acyclic species can be represented as a tree rooted at a unique point. A tree is a directed graph where one node plays the special role of a root. A similar approach has been implemented in the algorithms developed by Chevalier et al.[5] and Blurock.[6] In these systems external graph representation of species has been implemented. In Blurock's previous work, a concept of reaction pattern was introduced. The matrix representation system was developed in Klein's group and further enhanced in Broadbeld's, Dean's and Green's groups.[7,8] In this approach a molecule is represented as a square matrix (adjacency matrix), and chemical reactions as matrix transformation. Since different matrices can be used to represent a given molecule depending on the atom used as the reference point (the root), a matrix representation is not canonical. Molecular isomorphism is established by direct comparison of unique string codes. Come and co-workers[9−11] have developed a different approach used in the EXGAS program. In this approach, the molecule is externally represented by 1D string (linear chemical notation). Analogous to matrix representation this notation is not canonical. Therefore, a canonical form of molecular trees must be established.[10] Chemical reactions are then performed on the level of canonical molecular trees. Another 1D-string system has been developed by Ranzi et al.[13] These approaches have been applied successfully in many combustion systems.

2. A GRAPH THEORY BASED MECHANISM GENERATOR

We have developed an automatic mechanism-generating tool based on a reaction class approach and graph theory to incorporate new advances in kinetic theory for predicting thermal rate constants from first principles. We elected to
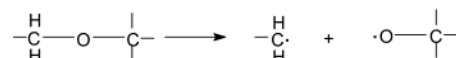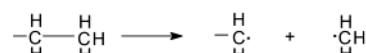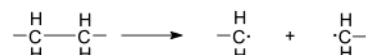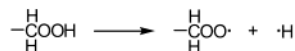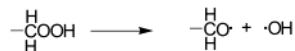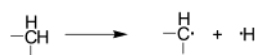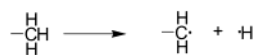
* Corresponding authors phone: (011)48857470113; e-mail: artur@mercury.hec.utah.edu (A.R.) and phone: (801)581-4301; e-mail: truong@chemistry.utah.edu (T.N.T.).
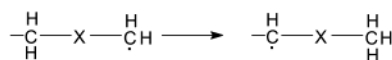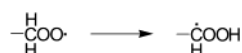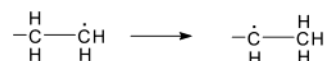† University of Utah.
‡ University at Bialystok.

Chart 1. List of Reaction Patterns Currently Available in the Program

Unimolecular reactions:

1. Bond fission.

(1a)

(2a)

(3a)

(4a)

(5a)

(6a)

(7a)

2. Rearrangement reactions:

a) Shifts: Any 1-n (n=2,3,...) C-C, C-O, O-C, O-O hydrogen shifts reactions, for example:

(8a)

(9a)

(10a)

(11a)

(12a)

where X – any C,O or N atom

(13a)

b) Others, for example cyclic ether formation:

(14a)

Bimolecular reactions:

1. Neutral – radical meta-thesis:

(15a)

(16a)

where R – any radical

(17a)

(18a)

(19a)

2. Neutral – neutral:

(20a)

(21a)
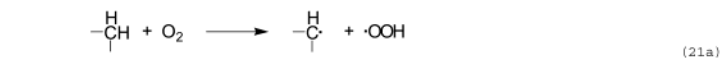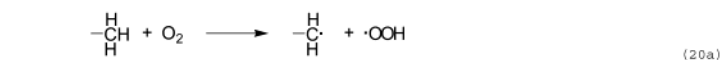
3. Radical – radical (recombination and disproportionation):

(22a)

(23a)

(24a)
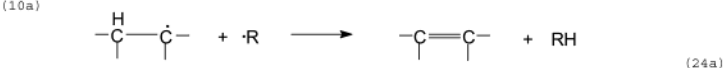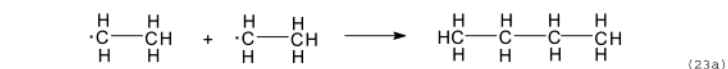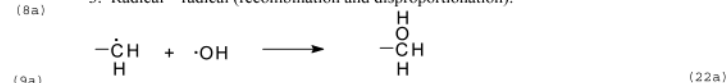
Others:

Double bonds are automatically created from diradicals:
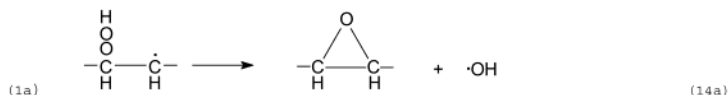
(25a)
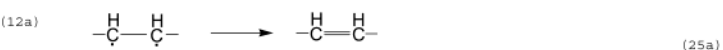
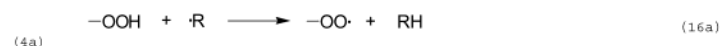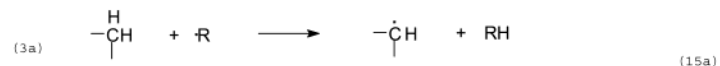The program automatically performs double bond migration:

(26a)

employ and extend a chemical graph theory approach used earlier by Blurock.[6] In particular, the present program called COMGEN (**CO**mplex **M**echanism **GEN**erator) uses chemical graphs in all the stages of the mechanism generation process for both external and internal representations. In this approach molecular species are naturally represented by a graph, identified by its vertices and edges where vertices are atoms in the molecule and edges are the bonds. Further, one can put additional molecular information beyond the connectivity, such as atomic charges, valences, bond types, etc., into the graph representation by using weighted vertices and edges. Such information is particularly useful to identify active sites in the reactant species, thus eliminating the possibility of generating unimportant product species a priori. The mechanism is generated using a set of reaction patterns (sub-graphs) that are the internal representations for given classes of reactions. Molecular species are internally represented by a connectivity table and uniquely identified by their molecular identification number, which is a set of topological
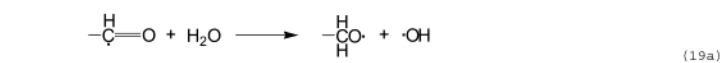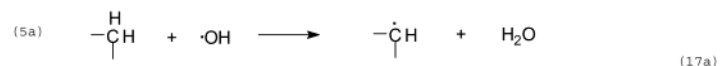
indices. Moreover, the use of reaction patterns to automatically generate reaction schemes can be very useful in the automatic reduction (lumping) schemes. The program is designed to generate all possible reactions in the reaction mechanism for given initial species and a set of reaction patterns.

2.1. Reaction Pattern. The key to the mechanism generation is a reaction pattern. The word "pattern" emphasizes that when the reaction pattern is applied, the pattern (substructure) of the reaction is matched with the same pattern in the molecule. The application of a reaction pattern on a molecule results in a specific reaction defined by such a pattern. Patterns are internally represented as chemical graphs and externally coded as 1D strings (we will deal with this in the following section). Some examples of reaction patterns already tested can be found in Chart 1. As one can see, every reaction pattern graph consists of two to four chemical graphs. In these figures, symbols "-" connected to only one atom represent so-called "free valences". The same
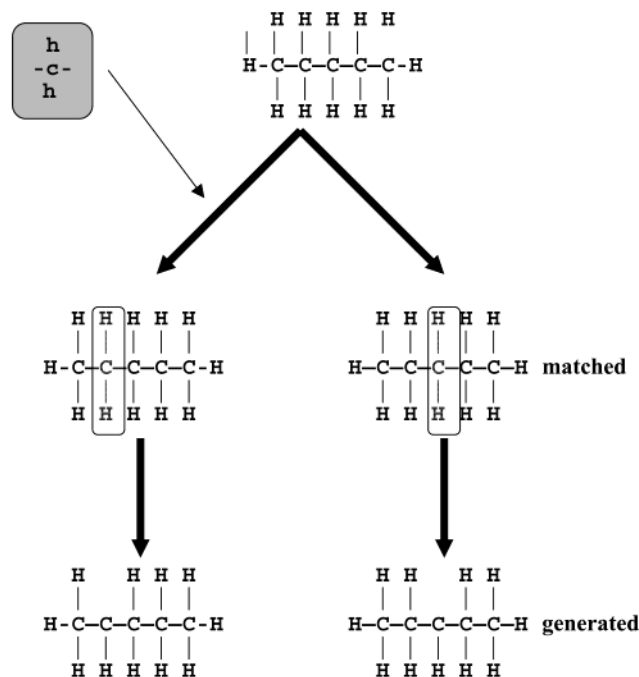
Figure 1. The process of application of the pattern to the molecule. The reaction on the left has multiplicity 4 (four hydrogen atom could be removed to have the same effect), whereas the right one has only 2.

symbols connected between two atoms represent single bonds whereas "=" symbols for double bonds. An example of a simple pattern graph is given below:
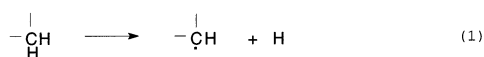
This pattern represents C−H bond dissociation from a secondary carbon -CH$_2$-. The "free valence" "-" symbols in eq 1 denote that this carbon atom is connected to two and only two other carbon atoms. Therefore, this pattern matches only secondary carbon atoms. If there were only one free valence, the pattern would match only primary carbon, such as patterns 1a, 3−4a from Chart 1. The generation of the new molecule by pattern application is done in two steps: (1) recognition of the pattern (substructure) in the molecule and (2) application of the pattern. The overall scheme for application of the pattern above is presented in Figure 1. All of the patterns, except radical recombination, alkenes formation from diradicals, and resonance reactions, are soft-coded, i.e., provided in the input file. Users have the option to neglect the hard-coded reaction classes. Because the pattern is represented as graph, new patterns can be easily created.
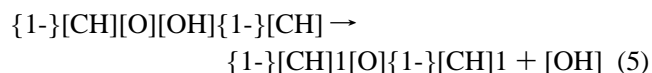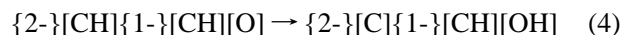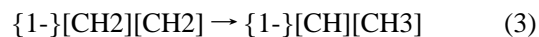
2.2. External Representation of Patterns and Species. The graphs for molecular species and reaction patterns used by the program are externally (in input and output files) represented by 1D linear chemical notation. There are several notation systems available in the literature (see, for example, refs 14 and 15). We used SMILES notation[15] due to its extensive use (to our knowledge, COMGEN is the first mechanism generator using this notation). SMILES syntax rules are widely available, for example from the developer website (www.daylight.com), thus they will not be presented here. In this notation all molecular features, even so subtle as optical isomerism, can be easily expressed. We have

introduced an extension to the SMILES standard in order to represent reaction pattern graphs. The "free valence" symbols defined in the previous section are represented as additional SMILES "letters" enclosed in "{" parentheses and are placed on the left-hand side of the atom with free valences. As an example, a reaction pattern from eq 1 is represented in the input file as

$$\{2\text{-}\}[CH2] \rightarrow \{2\text{-}\}[CH] + H \qquad (2)$$

Reaction patterns can contain "multiple bond" free valences and/or free valences connected to more than one atom. Equations (3) and (4) below show more complicated examples in the external notation. Shown below are patterns (8a), (10a) and (14a) from Chart 1.

$$\{1\text{-}\}[CH2][CH2] \rightarrow \{1\text{-}\}[CH][CH3] \qquad (3)$$

$$\{2\text{-}\}[CH]\{1\text{-}\}[CH][O] \rightarrow \{2\text{-}\}[C]\{1\text{-}\}[CH][OH] \qquad (4)$$

$$\{1\text{-}\}[CH][O][OH]\{1\text{-}\}[CH] \rightarrow$$
$$\{1\text{-}\}[CH]1[O]\{1\text{-}\}[CH]1 + [OH] \quad (5)$$

2.3. Pattern Recognition. Pattern recognition is one of the most important tasks in all the mechanism generation procedures. This process is looking for the atom-to-atom correspondence between the sub-graph of a reaction pattern and a given molecule. It produces all matches when a particular pattern is mapped on a given molecule. The program relies on a graph-theory based molecular tree isomorphism algorithm to find the matching positions in the molecule (see Figure 1, with more than one match when the pattern is applied). Due to the symmetry of the molecule, some patterns can have multiple matches that yield the same product. We define a "reaction multiplicity" as the number of possible symmetry-related reactions that yield the same product(s) for a given reaction pattern. For example, in Figure 1 C−H bond dissociation can occur on any of the four equivalent C−H bonds of the C2 and C4 secondary carbons to give the same products and thus its multiplicity is equal to 4. The other possible reaction at the C3 carbon has the reaction multiplicity of 2. Internally, each pattern or molecular species is represented by two sets of molecular trees starting from each carbon atom in the pattern or in the molecular species; the first set consisting of atomic symbols (representation of vertices) and the second of bond symbols (representation of edges). Examples of such trees and their computer representations are presented in Figure 2.

The pattern recognition procedure is illustrated in Figure 3. Four sets of molecular trees are created; two correspond to the molecule and two to the pattern. The pattern matches the molecule site if and only if there is a one-to-one correspondence between both trees from the molecular species and from the pattern. The procedure starts from splitting each tree into separate branches. The number of branches in the pattern tree is equal to numbers of atoms ending the tree (last leaves). The branches from the molecular tree are trimmed to have the same length if they are longer than branches from the pattern tree. In Figure 3, the atoms remaining after trimming of the molecular tree are marked in bold. Having completed the split, one can compare the two sets of branches. Because free valence symbol "-" ending a pattern branch is equivalent to a single bond to a carbon

AUTOMATED MECHANISM GENERATION

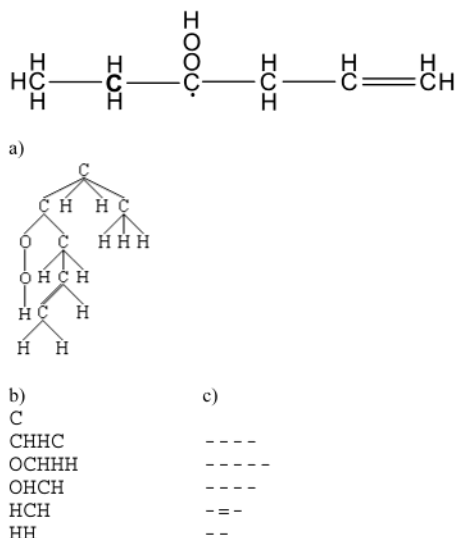*J. Chem. Inf. Comput. Sci., Vol. 43, No. 1, 2003* **39**

Figure 2. An example of a molecular tree with the bold carbon atom designated as the root. The molecular tree (a) is represented in computer memory as two character arrays: atomic tree (b) and bond tree (c).
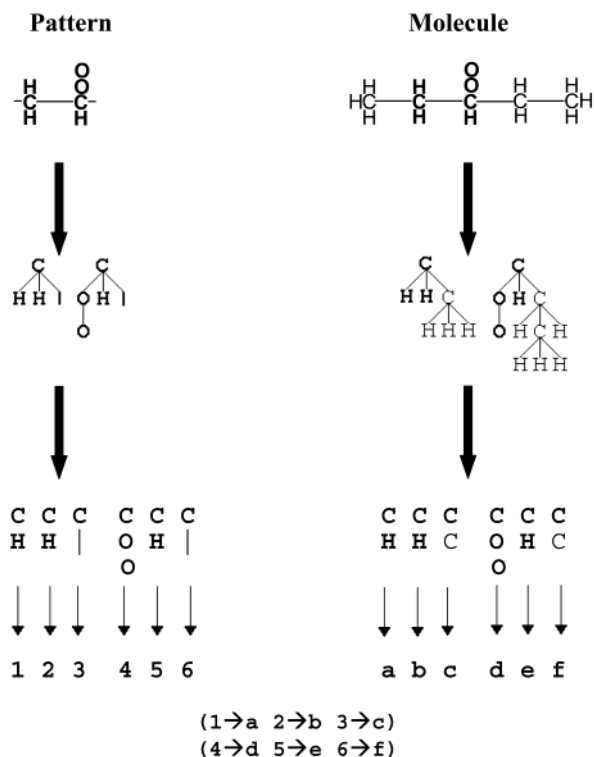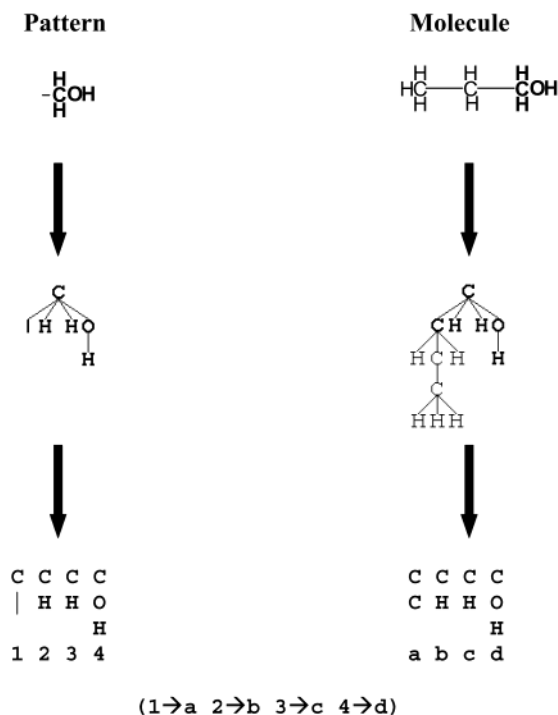

Figure 3. Substructure recognition of patterns containing one carbon atom. Due to the fact that in both structures only single bonds are present, bond trees are not listed for the sake of simplicity.

atom in the molecular tree, the pattern branch (1) in Figure 3 is equivalent only to branch (a) of the molecular tree. If there were two (three) free valences in the pattern tree, two (three) non-hydrogen atoms ending the molecular branches would be required to match the pattern. The procedure is more complex if the pattern involves more than one carbon atom, although the main idea is still the same. The program recognizes each of these carbons separately. Pattern (13a) from Chart 1 can serve as an example of such a procedure. This pattern represents 1,4 hydrogen shift from a secondary carbon to a peroxy oxygen anion (see Figure 4). To apply this pattern, two carbons must be matched. In this case, eight trees are created, four corresponding to the pattern and four

Figure 4. Substructure recognition of patterns containing two carbon atoms.

to the molecule. Branches from the pattern trees going through the carbon next to the root cut are removed during the tree creation process, as are the trees corresponding to the molecular species. Branches going through the second matched carbon atom next to the tree root are removed as well. The parts of trees remaining after these cuts are then compared in the same manner as illustrated in Figure 3. The pattern matches the molecule if and only if all corresponding pairs of trees are identical. One can easily see that there are two possible ways the pattern can be fitted to the molecule in Figure 4. Only one of the matches is marked by bold characters (C2 and C3 carbons in the molecule), whereas in the second match there are C3 and C4 carbons. Thus the reaction multiplicity in this case is equal to 2.

2.4. Reaction Pattern Application. The goal of pattern application procedure is to create new species and reactions. The technical details of the application process depend on the pattern. A simple reaction pattern involves the replacement of a substructure of the molecule by another substructure such as the reaction types 1−4a, 8−10a and 13−22a in Chart 1. Generally speaking, when there is only one reactant and product (rearrangement reactions) or when one of the two reactants and products are simple molecules or radicals such as $H, O, OH, OOH, H_2, O_2, H_2O, CO_2$, etc., substructure recognition can be done for only one reactant, and all products can be generated from only one substitution. For more complicated reaction types, specific procedures are needed to describe their applications. For example, when performing -C−C- or -C-OC- bond-fission reaction (patterns 5−7a), the program divides molecules with recognized substructures by removing the bond between two carbons or carbon and oxygen. Two new molecular species are then created. Similarly, the reverse direction of such reactions, i.e., radical recombination (22−23a), can also be done. A

different procedure is applied to meta-thesis reactions with two complex radicals (15−16a). In this case, two substructures have to be recognized. The program can also automatically convert created diradicals into alkene, as in the pattern 25a, and automatically perform bond migration in resonance-stabilized radical species, as in the pattern 26a. Currently, almost all reaction classes available in the literature[1,5,6,8] were tested. In fact, in principle there are no restrictions in the number of patterns available. One only needs to define such a pattern and then test it to see if a specific pattern application procedure is required.

2.5. Molecular Isomorphism. Molecular isomorphism is a procedure to determine the uniqueness of species generated. This procedure is needed to check if a newly generated species or reaction has previously been generated. If not generated previously such species/reactions will be added to the list of new species/reactions. The molecular isomorphism procedure, with proper canonicity algorithm, should be reliable, efficient, and provide simple interface for exchange of molecular data with other modules for obtaining thermodynamic properties of molecular species or kinetic properties of a given reaction.

Algorithms implemented in previous generators, where technical details are available, are based on the canonical forms of molecular trees[10] or unique strings associated with each species based on molecular trees or biconnected components (bicomps).[12] Some of these algorithms, like that described in details in ref 10 are dependent on the class of compound (there are different algorithms for cyclic and acyclic species). In this study, we decided to use the concept of molecular identification numbers (Molecular ID). Molecular ID consists of a set of real numbers (topological indices) associated with a given species (molecular graph). Two molecules are then equal if and only if all corresponding indices in both sets are equal. This algorithm is independent of the class of compound; it allows both cyclic and acyclic species to be handled equally well. Currently, this set consists of the following indices: Randic valence connectivity χ_n indices of all possible orders n (from 0th to (k-1)th) when k is a number of heavy atoms in molecule;[16,17] Wiener index;[18] Balaban index;[19,20] Schulz index (MTI);[17,21] determinant of (distance + adjacency) matrix (TI);[17,21] and WID index.[17] Randic valence connectivity indices account for heteroatoms (during calculation, a distinction between carbon and non-carbon non-hydrogen atoms such as oxygen or nitrogen is made), multiple bonds, and atomic valences. Other indices are based on the distance and adjacency matrices. These matrices have also been extended to hetero-systems by replacing the elements of the adjacency matrix and distance matrix corresponding to heteroatoms and multiple bonds with the values incorporating the corrections due to the changes induced when a carbon atom is replaced by a heteroatom. The modified matrices are calculated according to rules presented by Nikolic et al.[22] Such matrices are then used to obtain values of appropriate indices.

Note that the present approach has the advantage that the set of indices can be used to obtain a number of molecular or even chemical reaction properties by QSAR/QSPR relationships.[17−20] We are not using this feature in this work, however.

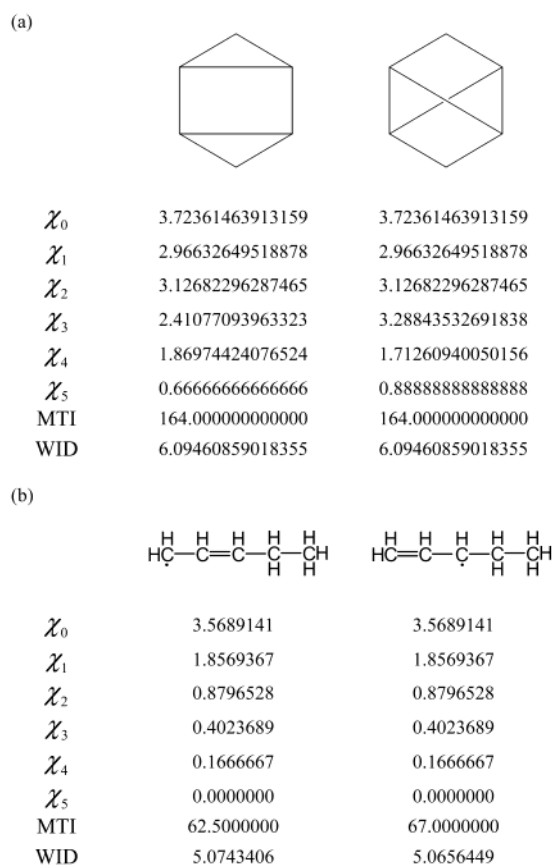The main drawback of a given topological index when compared to other methods lies in the problem of handling

Figure 5. Sets of selected topological indices for two pairs of very similar molecules.

a set of real numbers associated with each species generated. It is obvious that canonical strings (like in ref 12) are much easier and convenient to store and handle. Another possible weakness of the described approach is its discriminative power (possible degeneracy, e.g. two or more isomers may have the same topological index). By taking into account the heterogeneity of vertices (different atom types) and edges (different bonding types), such degeneracy can be effectively removed. Since different indices describe different characteristics of a given graph, the possibility for the set of indices chosen above to all have degenerate values is zero. This has been confirmed during our trial runs where thousands of species were generated and no two identical sets of indices for different species were found. In most cases, indices based on distance matrix are sufficient to distinguish between different species graphs. Among these, the WID (Weighted ID number) is believed to be the most discriminative.[17] For example, this index was tested for degeneracy on the fields of 1 346 024 alkane trees and no degenerate pair was found.[17] It fails in the very rare cases of cyclic species with multiple connections between atoms within a given cycle. Instead, Randic indices are able to make the distinction between such molecules. It is possible to find graphs possessing the same values of Randic indices, as well. It is true for some species, in particular that differ only by double bond migration. An interesting and illustrative example is shown in Figure 5. We have two pairs of very similar, almost identical, molecules together with corresponding sets of selected indices. Figure 5a shows two isomers that possess the same values of all distance-matrix based indices, including WID. Only three Randic indices are not the same. As mentioned,

AUTOMATED MECHANISM GENERATION

J. Chem. Inf. Comput. Sci., Vol. 43, No. 1, 2003 **41**

such situations are extremely rare. The situation reverses when one considers two molecules that differ only in the location of a double bond as shown in Figure 5b. In such a case only distance matrix-based indices are usable. It is obvious that in both instances, indices that do not account for different bonding types, valences, etc., would be the same. In fact, we found that to properly identify any molecular graph it is enough to calculate only Randic indices and one highly discriminative distance-matrix based index, e.g. Balaban or WID index. Consequently, the molecular ID provides a rather efficient approach to identify new species in a large set.

2.6. Thermochemistry. To use the generated mechanism in simulations of combustion systems, it is necessary to associate each molecular species and its thermodynamic properties, namely heat of formation, entropy and heat capacity, H_f^0, S°, C_p^0, respectively. We compute these properties by using the Benson group additivity approach.[23] This method allows one to compute molecules and radicals at any temperature between 300 and 1500 K. The main idea is to treat a molecular species as a sum of chemical groups. Each group makes its own contribution to the total thermodynamic quantities. A special procedure has been proposed for the case of radicals.[23] Furthermore, correcting for the symmetry of the molecule must also be included. The overall process of obtaining thermodynamic properties for a given molecule consists of three steps: (1) determination of the symmetry number of the molecular species; (2) recognition of group composition; and (3) calculating thermodynamic properties from the group additivity formula.

The first task is the most difficult and intricate one. The total symmetry number is a product of the external and internal numbers:[24]

$$\sigma = \sigma_{\text{external}} * \sigma_{\text{internal}}$$

Two publications dealing with calculating σ_{ext} have been published in the last 10 years.[24,25] In fact, methods described in both these publications are similar. The main difference lies in the manner of determining the molecular focus (symmetry center). In this work the symmetry center is found by using the method similar to that described in ref 24 and applied in the THERGAS software.[26] There are three possible symmetry centers,[24] namely a bond, an atom, or a ring and a different procedure is used to handle each case. When the focus is located on a bond then four bond trees are created from both atoms connected by the focus. The atoms lying on the "other side" of the focus are excluded from each tree. The rotational symmetries of each branch are determined. Both tree branches are then compared in a manner described in section 2.3. If they are identical, then $\sigma_{\text{ext}} = 2$ (branch rotational symmetry). If not then, $\sigma_{\text{ext}} = 1$. When the focus is an atom the numbers of trees are identical to the numbers of atoms directly connected to the focus. Similar to the former case, the rotational symmetry number of each branch and the number of branches identical (isomorphic) to each one in the set are determined. The procedure is identical to that in ref 24. When the focus is a ring, an appropriate symmetry correction according to the type of the ring is applied. Currently, the only ring available in the program is an epoxy ring. In such a case, σ is not equal 1 only when both branches started from the focus are identical (then

$\sigma_{\text{ext}} = 2$). Having calculated σ_{ext} the program determines σ_{int} for internal rotations in a manner described in detail in ref 23. An appropriate ring, optical isomers, gauche corrections, etc., are also taken into account.

Free radicals are treated using free radical group additivity for S° and C_p^0. Evaluation of H_f^0 is, like in THERGAS[26] and THERM[27] programs, based on the reaction

$$HR \rightarrow H \cdot + R \cdot$$

whereas a free radical group contribution method is used to calculate S° and C_p^0. The thermochemistry parameters obtained this way are stored as 14-coefficients polymonials in CHEMKIN format. The set of seven coefficients in range 300–1000 K is calculated with group additivity values. The extrapolation of C_p^0 at temperatures above 1500 K, the maximum temperature for Benson's method, was achieved in the following way. C_p^0 (5000 K) was approximated by $C_p^0(\infty)$, computed as value of classical limit. Between 1500 and 5000 K a Yuan-Mok relationship was used.[28]

For species that do not have group additivity parameters, one can obtain thermodynamic properties directly from quantum chemistry calculations. The program can create input files for each molecule generated. The 3D geometry is generated from data contained in the connectivity table.

2.7. Kinetics. At present, the same generic rate constant (set of Arrhenius parameters) is assigned to each pattern (reaction class). The multiplicity of each reaction is taken into account. The values of rate constants are taken from literature, such as refs 8 and 29–31. This will be changed soon by implementation of RC-TST (Reaction Class Transition State Theory).[32] This theory enables one to calculate the rate constant "*on the fly*", without finding the transition state. In fact, all the parameters needed to determine rate constant can be easily computed or estimated for each species. Currently, this theory is developed for only one reaction class (H abstraction),[33] but work is in progress on other classes. Moreover, the kinetic database populated with experimental and theoretically calculated rate constants is also under development.

2.8. Program Logic. In COMGEN the generation of a mechanism is done by iteratively applying a set of reaction patterns to a set of molecules. Successive iterations that consider all possible reactions between starting reactants, intermediates, and products builds a list of possible reactions (reaction network) according to the reaction patterns provided in the input file. Sometimes it is advisable to filter some unwanted reaction and/or species a priori, during the full mechanism generation. Such a procedure can be used to initiate preliminary mechanism simplification, before the specialized simplification tools, like CHEMKIN or KINALC programs, are applied. It is possible to limit the number of applications of a given pattern in the program. Moreover, the programmer can request species generated by a particular pattern not to be "touched" by other patterns. Another kind of restriction available is the possibility of setting the maximum carbon chain length of species created in radical recombination processes. This length is set up in the input file. Because no generic rules can be derived for the generation of the reactions involving radicals and molecules containing up to two carbon atoms this set of reactions, called C0–C2 reaction base, should be written down. In COMGEN,
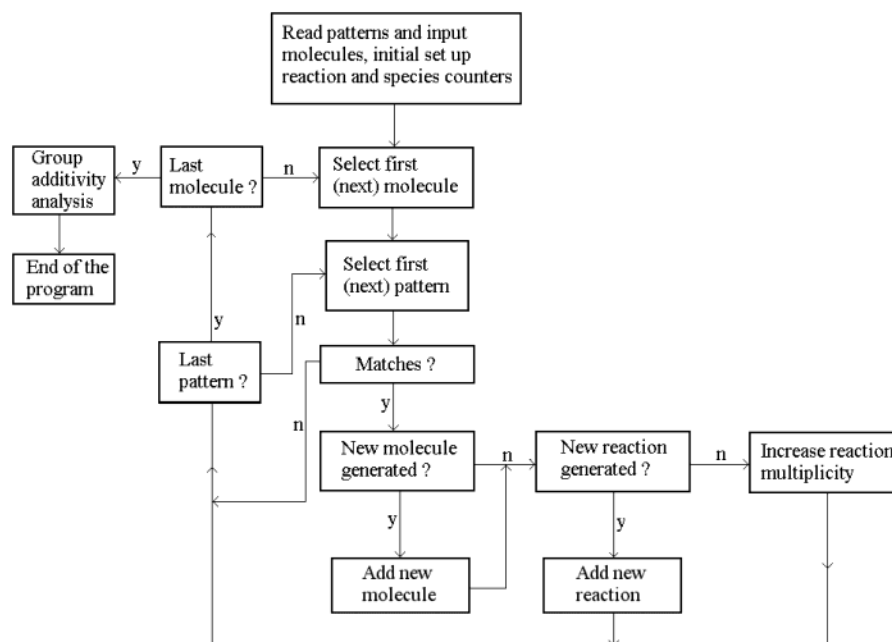
Figure 6. Program algorithm logic.

the C0−C2 reaction base developed by Come and co-workers[34] was used. This set consists of 426 reactions and 52 species. Of course, all the reactions already included in the C0−C2 reaction base are not written again in the comprehensive mechanism generated by the program.

The details of the algorithm are outlined in Figure 6. The reaction patterns are applied to all molecules until no new molecules are formed. All patterns are applied to the molecule at the same time. The program then goes to the next species.

3. EXPERIMENTS WITH COMGEN

The program is currently able to generate a complete mechanism of combustion of any acyclic hydrocarbon, alcohol, aldehyde, etc., and cyclic esters. As described in section 2.6, the mechanism is provided with complete (of course in the range of group additivity data available) set of thermochemical data. The output data file generated by the program is in CHEMKIN format. It consists of elements, species, thermochemistry data as 14 coefficients polymionals and kinetic data and can be used directly by CHEMKIN without any further modification. Moreover, molecular and reaction data are available in other formats, for example SMILES strings. Such a construction enables easy interfacing output results with other programs, like CHEMKIN, semiempirical or ab initio packages (like MOPAC) and 3D editors.

Several test runs dealing with combustion of propane have been performed. The input file consists of 23 reaction patterns with some application restrictions. Some (but not all) of these patterns are presented in Chart 1. Moreover, hard-coded patterns such as radical recombination are also used. The oxidation scheme similar to that proposed by Ranzi and co-workers[29] where the very initial formation of alkyl radicals takes place through the bimolecular chain initiation reactions with oxygen (patterns 20−21a in Chart 1) has been implemented.

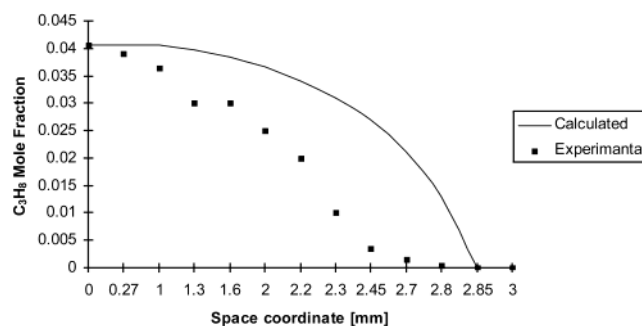To maintain reasonable number of resulting species and reactions, some restrictions described in section 2.8 were

Figure 7. Comparison between calculated and measured C_3H_8 mole fraction as function of space coordinate in premixed stoichiometric air-propane flame at the pressure $p = 0.25$ atm.

applied. For example the maximum number of carbon atoms in the recombination products was restricted to 3. Moreover, the number of meta-thesis reactions was limited by using the concept of β, μ, $\beta\mu$ and Y (resonance-stabilized) free radicals (see, for example, ref 11). These restrictions enabled us to keep the comprehensive mechanism at a manageable size. As a result, a mechanism consisting of 967 reactions and 194 species was obtained (these amounts include reactions and species from C0−C2 reaction base). This mechanism has been used to perform simulations with the CHEMKIN II suite of programs.[35] The stoichiometric propane-air laminar, premixed flame at pressure 0.25 atm was investigated using the PREMIX code. The temperature profile measured in ref 36 was used. These calculations are compared to experimental values obtained by Fristrom and Westenberg.[36] Figures 7 and 8 present, respectively, the mole fraction of propane and carbon dioxide in the flame as a function of the space coordinate. In both cases, the agreement obtained is rather satisfactory. Calculations of flame speed in freely propagating adiabatic flame ($p = 1$ atm) were also carried out. Results are presented in Figure 9. They are compared to measurements[37] and flame speeds obtained with GRI mechanism 3.0.[38] In this case, comparison to the experiment is less satisfactory than in the previous case. It is probably caused by inappropriate values of transport
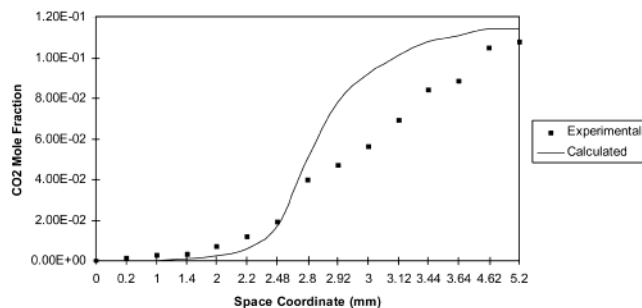
AUTOMATED MECHANISM GENERATION

J. Chem. Inf. Comput. Sci., Vol. 43, No. 1, 2003 **43**

Figure 8. Comparison between calculated and measured CO_2 mole fraction as function of space coordinate in premixed stoichiometric air-propane flame at the pressure $p = 0.25$ atm.
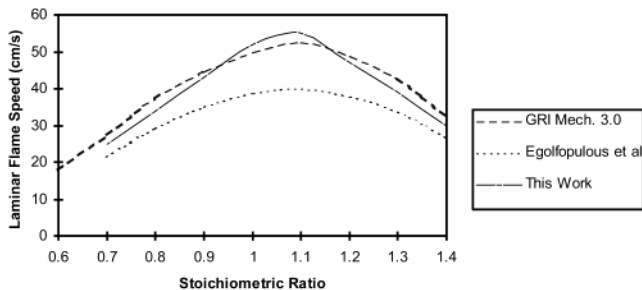

Figure 9. Flame velocities as function of stoichiometric ratio in premixed air-propane flame at the pressure $p = 1$ atm.

parameter coefficients used. They are known for only about 80 species from almost 200 involved in the mechanism. The program only roughly estimates the unknown coefficients. For the same reason, velocities obtained with COMGEN are almost the same as those resulting from the GRI mechanism, which consists of only a minimal set of propane kinetics (two species and 15 reactions).

4. CONCLUSION AND FUTURE DIRECTION

In this study, we present idea, algorithm and simple applications of the complex mechanism generator. Its main advantage is that the user can characterize the entire complex mechanism in terms of a small set of reaction patterns. Expanding these to the full set of reactions is done automatically. The mechanism is provided with a rather complete set of thermodynamic and kinetic data in a form ready for further kinetic analysis. Moreover, data files for further quantum chemistry analysis of species generated could also be provided. However, some improvements are needed for the system to quantitatively describe practical combustion systems. In particular, these are as follows:

1. Implementation of RC-TST theory and program database with data mining capabilities in order to obtain more accurate rate constants for each reaction generated.

2. Implementation of techniques that enable reduction of mechanism size without losing too much important information, like a rate-based scheme[8] or lumping.[11]

3. Calculation of the heat capacity and entropy calculation for radicals as bond dissociation values. This would significantly broader the spectrum of species for which thermodynamic data are available.

4. Implementation of QSAR/QSPR relationships to calculate some parameters that cannot be calculated in a different way, such as transport properties.

REFERENCES AND NOTES

(1) Glaude, P. A.; Battin-Leclerc, F.; Fournet, R.; Warth, V.; Come, G. M.; Scacchi, G. Construction and simplification of a model for the oxidation of alkanes. *Combustion Flame* **2000**, *122*, 451−462.

(2) De Witt, M. J.; Dooling, D. J.; Broadbelt, L. J. Computer Generation of Reaction Mechanisms Using Quantitative Rate Information: Application to Long-Chain Hydrocarbon Pyrolysis. *Ind. Eng. Res.* **2000**, *39*, 2228−2237.

(3) Tomlin, A. S.; Turanyi, T.; Pilling, M. J. In: *Low-Temperature Combustion and Auto-Ignition*; Pilling, M. J., Ed.; Elsevier: Amsterdam, 1997; Vol. 35, Chapter 4, p 293.

(4) Chinnick, S. J.; Baulch, D. L.; Ayscough, P. B. An Expert System for Hydrocarbon Pyrolysis Reactions. *Chemometr. Intell. Lab. Syst.* **1988**, *5*, 39−52.

(5) Chevalier, C.; Warnatz, J.; Melenk, H. Automatic generation of reaction mechanisms for the description of the oxidation of higher hydrocarbons. *Bes. Buns. Phys. Chem.* **1990**, *94*, 1362−1367.

(6) Blurock, E. S. Reaction: System for Modeling Chemical Reactions. *J. Chem. Inf. Comput. Sci.* **1994**, *35*, 607−616.

(7) Broadbelt, L. J.; Stark, S. M.; Klein, M. T. Computer generated reaction networks: on-the-fly calculation of species properties using computational quantum chemistry. *Chem. Eng. Sci.* **1994**, *49*, 4991−5010.

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

(9) Glaude, P. A.; Warth, V.; Fournet, R.; Battin-Leclerc, F.; Come, G. M.; Scacchi, G. Modeling of *n*-heptane and isooctane gas-phase oxidation at low temperature by using computer-aided designed mechanism. *Bull. Soc. Chim. Belg.* **1997**, *6*, 343−348.

(10) Warth, V.; Battin-Leclerc, F.; Fournet, R.; Glaude, P. A.; Come G. M.; Scacchi, G. Computer based generation of reaction mechanisms for gas-phase oxidation. *Comput. Chem.* **2000**, *24*, 541−560.

(11) Warth V.; Stef N.; Glaude, P. A.; Battin-Leclerc, F.; Scacchi, G.; Come, G. M. Computer-aided derivation of gas-phase oxidation mechanisms: application to the modeling of the oxidation of *n*-butane. *Combust. Flame* **1998**, *114*, 81−102.

(12) Broadbelt, L. J.; Stark, S. M.; Klein, M. T. Computer generated reaction modelling: decomposition and encoding algorithms for determining species uniqueness. *Computers. Chem Eng.* **1995**, *20*, 113−129.

(13) Ranzi, E.; Sogaro A.; Gaffuri P.; Pennati G.; Westbrook C. K.; Pitz W. J. A new comprehensive reaction mechanism for combustion of hydrocarbon fuels. *Combust. Flame* **1994**, *99*, 201−211.

(14) Come, G. M.; Muller C. A linear chemical notation. *Comput. Chem.* **1984**, *8*, 233−237.

(15) Weinenger D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 31−36.

(16) Randíc, M. J. Characterization of molecular branching. *J. Am. Chem. Soc.* **1975**, *97*, 6609−6615.

(17) Trinajstic, N. *Chemical Graph Theory*; CRC Press: Boca Raton, 1992; Chapter 10, p 225.

(18) Wiener, H. Structural determination of paraffin boiling points. *J. Am. Chem. Soc.* **1947**, *69*, 17−20.

(19) Balaban, A. T. Topological indexes based on topological distances in molecular graphs. *Pure Appl. Chem.* **1983**, *55*, 199−206.

(20) Balaban, A. T. Applications of graph theory in chemistry. *J. Chem. Inf. Comput. Sci.* **1985**, *25*, 334−343.

(21) Schulz, H. P. Topological organic chemistry. 1. Graph theory and topological indices of alkanes. *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 227−228.

(22) Nikolic, S.; Trinajstic, N.; Michalic, Z. Molecular topological index: an extension to heterosystems. *J. Mater. Chem.* **1993**, *12*, 251−264.

(23) Benson, S. W. *Thermochemical Kinetics*, 2nd ed.; Wiley: New York, 1976; Chapter 2, p 19.

(24) Muller, C.; Scacchi, G.; Come G. M. A topological method for determining the external symmetry number of molecules. *Comput. Chem.* **1991**, *15*, 17−27.

(25) Walters, W. P.; Yalkowsky, S. H. ESCHER − A computer program for determination of external rotational symmetry numbers from molecular topology *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 1015−1017.

(26) Muller, C.; Michel, V.; Scacchi, G.; Come, G. M. THERGAS: a computer program for the evaluation of thermochemical data of molecules and free radicals in the gas phase. *J. Chim. Phys.* **1995**, 92, 1154−1178.

(27) Ritter, E. R.; Bozelli, J. W. THERM: thermodynamic property estimation for gas-phase radicals and molecules. *Int. J. Chem. Kinet.* **1991**, *23*, 767−778.

(28) Yuan, S. C.; Mok, Y. I. New Look at Heat Capacity Prediction. *Hydrocarbon Processing* **1968**, *47*, 133−136.

(29) Ranzi, E.; Faravelli, T.; Gaffuri, P.; Pennati., G. C.; Sogaro, A. A wide range modeling study of propane and *n*-butane oxidation. *Combust. Sci. Technol.* **1994**, *100*, 299−330.

(30) Warnatz, J. The mechanism of high-temperature combustion of propane and butane. *Combust. Sci. Technol.* **1983**, *14*, 177−200.

(31) Battin-Leclerc, F.; Glaude, P. A.; Warth, V.; Fournet, R.; Scacchi, G.; Come G. M. Computer tools for modeling the chemical phenomena related to combustion. *Chem. Eng. Sci.* **2000**, *55*, 2883−2893.

(32) Truong, T. N. Reaction class transition state theory: Hydrogen abstraction reactions by hydrogen atoms as test cases. *J. Chem. Phys.* **2000**, *113*, 4957−4964.

(33) Zhang, S.; Truong, T. N. Kinetics of hydrogen abstraction reaction class H + H−C(sp^3): An application of the reaction class transition state theory. *J. Phys. Chem. A*, in press.

(34) Barbe, P.; Battin-Leclerc, F.; Come G. M. Experimental and modeling study of methane and ethane oxidation between 773 and 1573K. *J. Chim. Phys.* **1995**, *92*, 1666−1692.

(35) Kee, R. J.; Rupley, F. M.; Miller, J. A. Sandia Laboratories Report, 1993; SAND 89-8000B.

(36) Fristrom, R. M.; Westenberg, A. A. Flame zone studies IV − microstructure and material transport in a laminar propane-air flame front. *Combust. Flame* **1957**, *1*, 217−228.

(37) Vagelopoulos, C. M.; Egolfopoulos, F. N. Direct experimental determination of laminar flame speeds. *27th (Int.) Symp. On Comb.* **1998**, 513−519.

(38) Smith, G. P.; Golden, D. M.; Frenklach, M.; Moriarty, N. W.; Eiteneer, B.; Goldenberg, M.; Bowman, C. T.; Hanson, R. K.; Song, S.; Gardiner, W. C.; Lissianski, V. V.; Qin, Z. GRI-MECH 3.0 http://www.me.berkeley.edu/gri_mech/.