Reaction: System for Modeling Chemical Reactions[†]

Edward S. Blurock

Research Institute for Symbolic Computation, Johannes Kepler Universität, Linz, Austria

A complex chemical process, such as combustion, can be modeled by a list of individual reactions, called a mechanism. *Reaction* is a software system for the management, manipulation, and generation of molecular and reaction data. An important aspect of the system is the automatic generation of a detailed mechanism given a set of reaction classes. The internal software representation of a reaction class, a set of reaction patterns, is the key to the generation process. In this paper, the generation of a mechanism is explained. The iterative application of the reaction patterns on sets of molecules is described and an example, involving the low temperature branching agent formation in combustion, for a set of hydrocarbons is shown. *Reaction* represents the first stage of automating the generation and further study of mechanisms.

1. INTRODUCTION

A complex chemical process, such as combustion, can be modeled with a list of individual reactions called a mechanism. A mechanism study proposes a set of reactions, runs a numerical calculation, and compares the result with experiment. As the physical processes to be modeled become more complicated so do the mechanisms, and, thus, it becomes increasingly difficult to accurately write down the corresponding set of reactions. For example, combustion mechanisms dealing with the *n*-alkanes are considerably simpler than those dealing with highly branched hydrocarbons. The reactions dealing with the *n*-alkanes, in a sense, only involve linear chains of primary and secondary carbons. However, those dealing with highly branched hydrocarbons involve combinations of primary, secondary, tertiary, and quaternary carbons in complex interactions. Since each mechanism can have hundreds (and maybe thousands) of individual reactions, writing down the mechanism can be tedious and error-prone. Moreover, with so many reactions, the structure within the mechanism is often concealed or lost. For these reasons it is important to develop automatic systems to accomplish this task.

Reaction is a system for the manipulation and management of molecular and reaction information. One feature of *Reaction* is the automatic generation of mechanisms stemming from one or more initial molecules. The output is compatible with standard packages for solving the associated system of differential equations. In *Reaction*, the key to mechanism generation is the reaction pattern.

1.1. Reaction Patterns. A mechanism is generated using a set of reaction patterns[1,2] which is the internal representation of a class of reactions. The word "pattern" is used to emphasize that when the reaction pattern is applied, the pattern (i.e., the substructure) of the reaction is matched with the same pattern within the molecule. The application of a reaction pattern on a molecule results in a specific reaction.

For example, one class of reactions is hydrogen abstraction from a carbon. The corresponding reaction pattern is
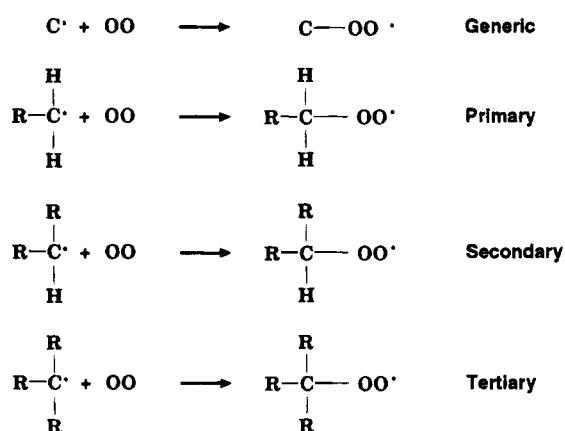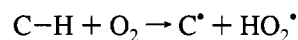
Figure 1. Shown are four representations of reaction patterns for hydrogen abstraction from carbon, *generic* (i.e., representing the entire class of simple hydrogen abstraction) and the three more specific representations, hydrogen abstraction from primary, secondary, and tertiary carbon centers.

$$C-H + O_2 \rightarrow C^{\bullet} + HO_2^{\bullet}$$

and the application of the hydrogen abstraction reaction on butane results in the reaction

$$C_4H_{10} + O_2 \rightarrow C_4H_9 + HO_2$$

The result of this application is actually two reactions: abstraction of a primary hydrogen and abstraction from a secondary carbon. Furthermore, due to the symmetry of the molecule, abstraction from the primary carbon can occur with any one of six hydrogens and abstraction from the secondary carbon can occur with one of four different carbons. This results in a "multiplicity" associated with each reaction.

A class of reactions is either represented by one reaction pattern, as above, or by a set of reaction patterns. For example, the reaction class addition of O_2 to a carbon radical could be represented as the single reaction pattern shown at the top of Figure 1. In this case, all carbon radicals would be treated the same, i.e., all reactions derived from this reaction pattern would have the same kinetic constants. However, if a distinction in kinetic constants is to be made between primary, secondary, and tertiary carbons, then the reaction class should be represented by a set of three reaction patterns (see Figure 1).

Hydrogen Abstraction

OO•   R• + O₂ ⇌ ROO•

OOH   **1,4 QOOH Formation**

OOH   R• + O₂ ⇌ ROO•

OO•

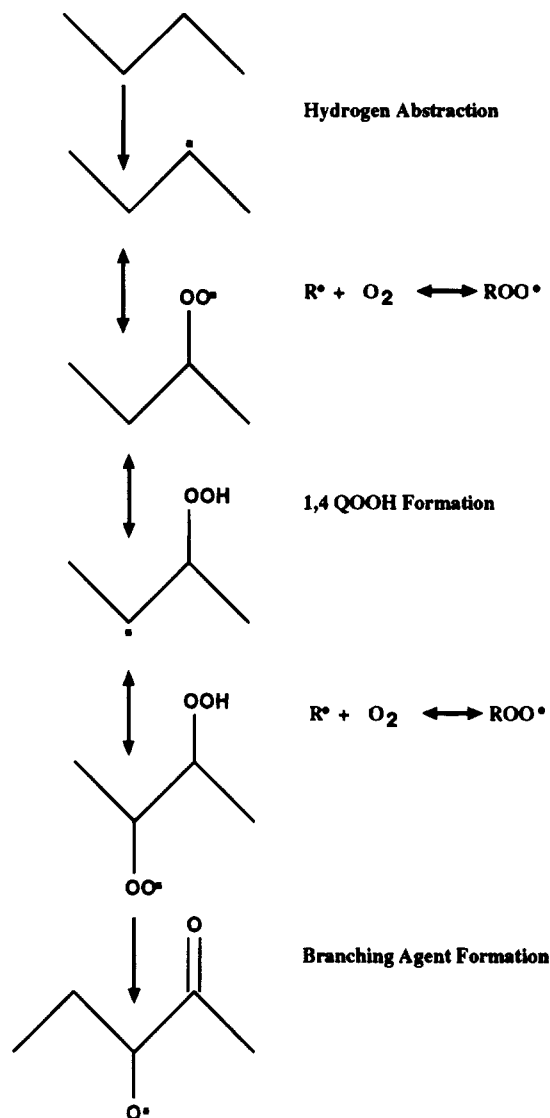O   **Branching Agent Formation**

O•

Figure 2. This is an example of the progression of reactions and intermediates to the branching agent. Double arrows show those reactions which were considered reversible.

Reaction

(1) [Initialization] The initial sets of reactants and reaction patterns are set up. These are chosen from the list of reactions and molecules available in the database of the system.

(2) [Apply Reactions]Each reaction pattern is, in turn, applied to the set of molecules in the active list. A set of products are produced and added to the total set of molecules in the system.

(3) [Iterate] The set of active molecules is replaced by a new list of those molecules produced in the last application of the reaction patterns. The reaction patterns are applied again, as above. This process continues until no new molecules are produced or a maximum number of iterations is reached. In the case of sequential operation, a new set of reaction patterns is applied in each iteration.

(4) [Kinetic Analysis]The system of reactions and molecules is then sent to a kinetic module to solve, numerically, the set of associated differential equations. The result is a plot of the changes in concentration of the species.

Figure 3. The overall algorithm of reaction mechanism generation and analysis.

The use of reaction patterns is widespread in reaction databank management and computer aided synthesis design.[3-5] The representation of reaction classes as rules (i.e., *if* these conditions hold, *then* perform these operations) has been used to generate combustion mechanisms.[6-10] In addition, the concept of reaction classes, whether represented as rules or as reaction patterns, is, in a sense, used within the concepts of "lumping" and reduced reaction schemes.[11-15] In this case, reaction classes are not used to generate the reaction schemes but to simplify them. It is hoped that the use of reaction patterns to automatically generate reaction schemes

will eventually be useful in the automatic reduction of those schemes, i.e., to automate the lumping procedures.

2. REPRESENTATION OF MOLECULES AND REACTIONS

The basic representation of the molecule is graphical, i.e., a set of nodes, the atoms, and a set of bonds connecting the nodes. This graphical description is expanded by additional atom and bond information (such as the results of semiempirical calculations, lone pair and aromatic determination, ring recognition, etc.). The representation of the molecule for use in mechanism generation has all the atomic information packed into a single integer. Only single graphical bonds exist between the atoms. Multiple bond information is contained on the atoms themselves. This is done to simplify the substructure search used in reaction pattern recognition. Conversion back to the multiple bond form does not present a problem.

A reaction is represented as the changes between the reactants and the products. The representation of the reaction pattern consists of the following parts:

Reactant Subgraphs	This is a set of graphs that are to be matched in the reactant molecules. These atoms and bonds changed within the course of the reaction. It should be noted that atoms not directly involved in the reaction can be included to more exactly characterize the reaction (as in the hydrogen abstraction example for primary, secondary, and tertiary carbons).
Product Subgraphs	This is the set of atoms and bonds (subgraphs) that result from the changes.
Valence Changes	This is a set of pairs: the atom valence of the reactant and the atom valence of the product. This represents the change in valence of the atom as a result of the reaction.
Unmatched Reactant Atoms	These are the atoms which appear on the reactant side of the reaction but not on the product side. This possibility is included because in some databases some atoms are assumed but not explicitly listed.
Unmatched Product Atoms	These are the atoms which appear on the product side of the reaction but not on the reactant side.
Bond Changes	These are the set of bonds that have changed within the course of the reaction. A bond is only listed here if the bond disappears completely. The double bond changing to a single bond is represented via a valence change and hence does not appear in this list.

The key step in the application of a reaction is the matching of the reaction pattern to atoms within the molecule. This is done by matching the individual reactant subgraphs to the molecules.
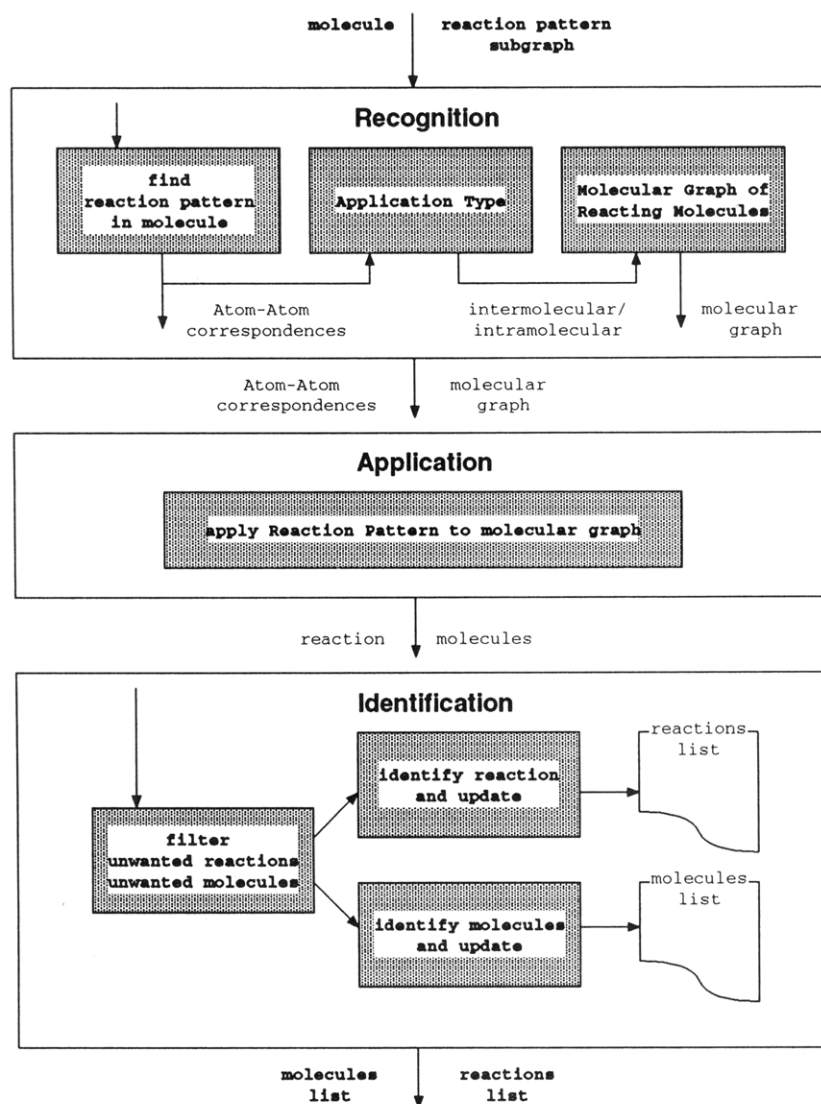
Figure 4. This gives the overall structure of the application of a single reaction pattern on a set of molecules.

Recognition(RxnPat, Mols)

Input: *RxnPat* the current reaction pattern to be used.
       Mols are the molecules on which the reaction pattern *RxnPat* is to be applied.

Output: *AtomCorrs* are the atom-to-atom correspondences between the subgraphs of the reaction pattern and the set of molecules.
       MolGraphs are the set of molecular graphs, each corresponding to a specific atom-to-atom correspondence in *AtomCorrs*.

(1) [Find the reaction pattern in the molecules] Find all combinations of how the reaction pattern fits into the set of molecules. The positions (there can be more than one) within the molecule where the reaction pattern subgraph can be applied are determined using a graphical comparison. Each correspondence is represented as a set of pairs, where the first element is the atom in the molecule and the second the atom in the reaction pattern.

(2) [Application Type] Based on the set of correspondences and the set of molecules involved the type of application (intermolecular between two different molecules, intermolecular between two of the same molecules and Intramolecular) is determined. The decision is based on whether there are conflicts between the subgraph atom-atom correspondences.

(3) [Molecular Graph of Reacting Molecules] Those molecules with which the subgraphs of the reaction pattern have matched are put together in one molecular graph.

Figure 5. This procedure finds the atom—atom correspondences between the reaction pattern and the set of molecules. It determines whether the reaction is to be inter- or intramolecular.

3. REACTION PATTERN APPLICATION

In *Reaction* the generation of a mechanism is done either by iteratively applying a set of reaction patterns to a set of molecules or by sequentially applying a list of reaction classes (where each class is represented by a set of reaction patterns) to a set of molecules. In the iterative application, the reaction patterns are applied to all the molecules until no new molecules are formed. In the sequential application each set of reaction patterns are applied to the molecules-

generated in the last step until no more reaction patterns are left in the sequence. The basic operation in both cases is the application of a set of reaction patterns to a set of molecules.

For example, suppose a simplified version of the low temperature reaction mechanism of combustion has been proposed and is to be applied to a single molecule. The proposed mechanism, shown in Figure 2, consists of a sequence of five reaction classes: hydrogen abstraction, addition to form peroxy radical, rearrangement, addition of a second oxygen and finally, and formation of the branching agent. The generated mechanism is the result of applying these reaction classes to a given molecule producing a set of specific reactions. The system proceeds by first performing the hydrogen abstraction on the molecule. A set of radicals are produced. The system then applies the second reaction class of the mechanism, the addition of oxygen, to this set of radicals to produce the peroxy radicals. When all five reaction classes are applied, the entire generated mechanism for the molecule is produced.

The overall algorithm is outlined in Figure 3. The details of the reaction mechanism generation are given in the next sections.

3.1. Procedure. One cycle of the generation procedure consists of three phases (see Figure 4):
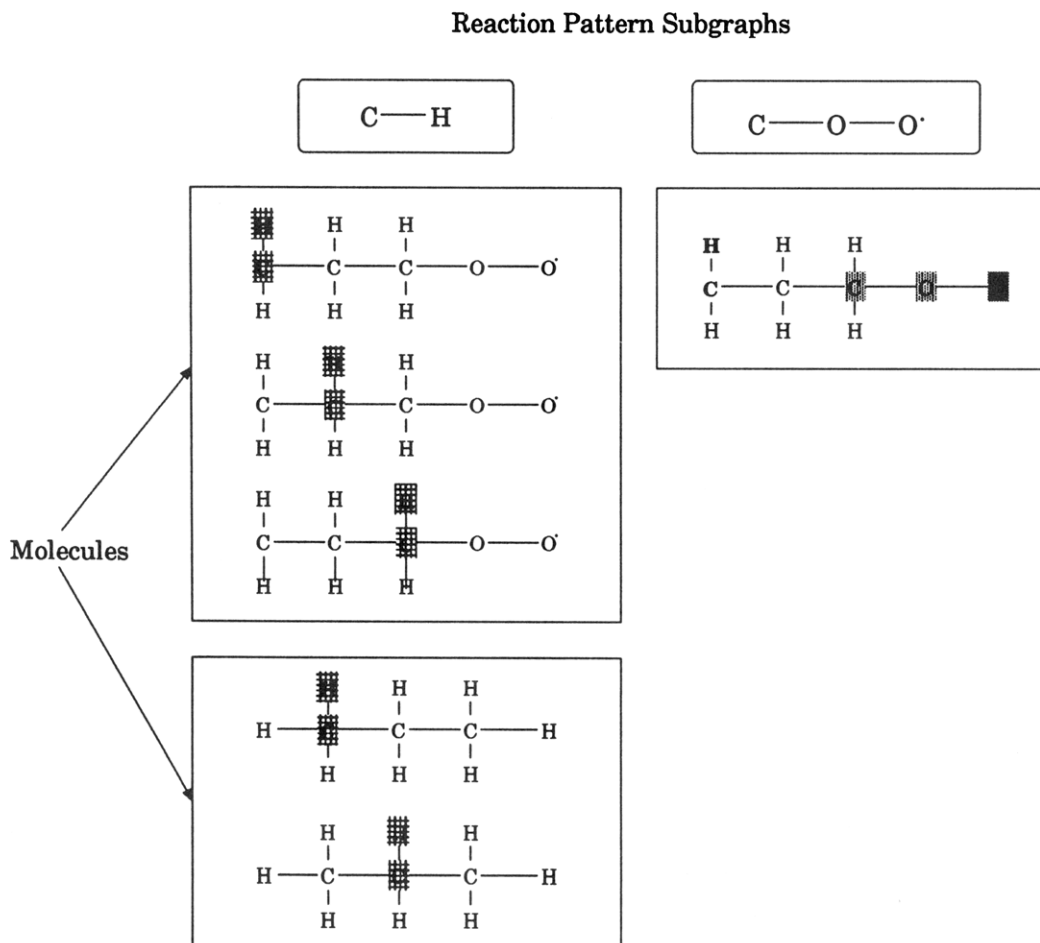
Reaction Pattern Subgraphs

Figure 6. This is an illustration of the matching of a reaction pattern made up of two patterns to two different molecules. Only the unique matches are shown.

Recognition　A list of correspondences (atom-to-atom matches) is made between the set of molecules and all reactant subgraphs in the set of reaction patterns.

Application　Using the correspondences, the valence and bond changes are applied to the molecules involved.

Identification　The list of molecules and reactions are checked against the already existing ones. If new, they are added to the total list.

3.1.1. Recognition. The recognition phase of the procedure finds correspondences between the reaction pattern and the set of molecules. The procedure basically finds how the reactant subgraphs making up the reaction pattern fit into the sets of molecules. It then applies the corresponding changes in bonds and valence. The algorithm is shown in Figure 5.

When the reaction pattern is made up of two separate reactant subgraphs, such as the case (shown in Figure 6) with peroxyl radical extracting a hydrogen from a carbon, the application of the reaction pattern can lead either to intramolecular reactions or intermolecular reactions. The first step is to find all the possible matches of the reactant subgraphs in the molecules. In the next step, each of the five carbon–hydrogen matches (left side of figure) is paired with the single peroxy match (right side of figure). These sets of pairs represent the possible applications of the reaction.

Depending on the pair, an intramolecular or an intermolecular reaction can be represented.

Figure 7 shows some examples of the application of such a reaction pattern in the recognition procedure. The reactant subgraphs are matched in each of the molecules (in the case of propane, there is no match of the peroxy group). Then, depending on the resulting pair, the type of application is determined and the result computed. In this case, the result can be one or two new molecules. The figure shows four illustrative cases (from top to bottom):

Intramolecular　Both reactant subgraphs are matched within the same molecule. There are no conflicts between the atom–atom correspondences so this could be an intramolecular reaction. It should be noted that this could also be treated as an intermolecular reaction (but this is not done).

Intermolecular　Both reactant subgraphs are matched within the same molecule. However, there are conflicts with the atom–atom correspondences. Thus, this must be an intermolecular reaction between two identical molecules.

No Reaction　One of the reactant subgraphs is matched, but the other did not match within the molecule. This means that there can be no inter- or intramolecular reactions with itself.
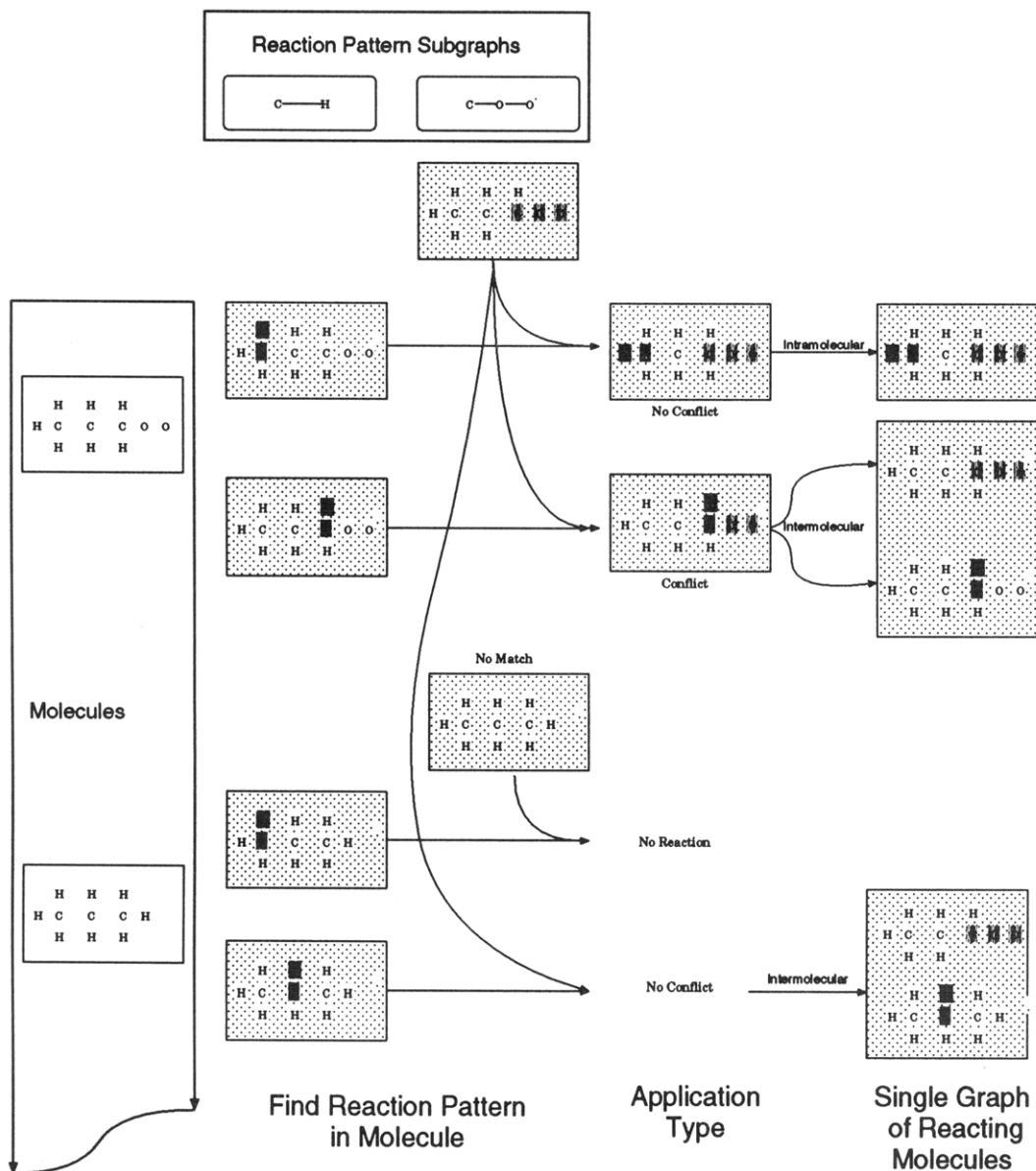
Figure 7. This is an illustration of the *Recognition* procedure with four different cases (see text for details).

Identification(MechRxns, MechMols)

Input: *MechRxns* are the new reactions.
          MechMols are the molecules within the new reactions.

(1) [Filter Unwanted Reactions and Molecules] Filter out unwanted reactions (*MechRxns*) and molecules (*MechMols*). Typical filters include putting a limit on the size of molecules produced and eliminating molecules with two or more radicals.

(2) [Identify Molecules and Update]Each molecule produced is checked against the list of molecules in the master molecule list. If the molecule is already present, then its identification number is noted. If the molecule *MechMol$_i$* is not in the list, then the molecule is added to the list and the new identification is noted.

(3) [Identify Reactions and Update] The ID's of the reaction are first updated. Then the reaction *MechRxn$_i$* is then compared with those already in the list. If it is unique, then it is added to the list. If it exists in the list already, then the count of this reaction is updated.

Figure 8. This algorithm filters out unwanted reactions and molecules and puts them into the master molecule and reaction lists.

Two Molecules   The first reactant subgraph matched the second molecule, and the second reactant subgraph matched the first molecule. This is an intermolecular reaction between two different molecules.

3.1.2. Application and Identification. Given an atom-to-atom matching of the reactant subgraphs of the reaction pattern to the molecule(s), the same bond and valence

```
NONE
1000
2010,2020,2030;Oxygen
3001,3002,3003
2010,2020,2030;Oxygen
6011,6021,6012,6022,6013,6023
```

Figure 9. The reaction class sequence used in this experiment is shown. The sequence consists of five steps. The NONE specifies that there are no molecules which should be present for all reactions. The two references to Oxygen indicate that Oxygen should be included along with the intermediates of the previous reactions. Each line represents a step in the sequence. The set of reaction patterns given on one line are applied together.

changes that occur within the reaction pattern (in going to the product subgraphs) are applied to the corresponding atoms in the molecule(s). The result is a new set of molecules.

The molecules and the reactions produced are then checked against the lists of existing molecules and reactions for duplication. If they are unique, they are added to theirrespective lists. If the reaction is not unique, then a multiplicity

612 *J. Chem. Inf. Comput. Sci., Vol. 35, No. 3, 1995*

BLUROCK

Table 1. Rate Constants Used in the Low Temperature Oxidation Mechanism

description	direction	A	E/R
($RH + O_2 \rightarrow R^{\cdot} + HO_2^{\cdot}$)			
initiation step			
1000	forward	6.022×10^{10}	2610
($R^{\cdot} + O_2 \rightarrow RO_2^{\cdot}$) primary			
2010	forward	5.42×10^9	0
2010	reverse	7.6×10^{16}	16358
($R^{\cdot} + O_2 \rightarrow RO_2^{\cdot}$) secondary			
2020	forward	5.42×10^9	0
2020	reverse	6.0×10^{15}	17080
($R^{\cdot} + O_2 \rightarrow RO_2^{\cdot}$) tertiary			
2030	forward	5.42×10^9	0
2030	reverse	2.8×10^{15}	15625
1,4-QOOH formation			
3001	forward	1.41×10^{12}	18643
3001	reverse	8.63×10^{10}	12000
1,5-QOOH formation			
3002	forward	1.76×10^{11}	14758
3002	reverse	3.33×10^9	8000
1,6-QOOH formation			
3003	forward	2.20×10^{10}	12617
3003	reverse	6.72×10^9	6170
1,4-branching agent (primary)			
6011	forward	1.41×10^{12}	16410
1,4-branching agent (secondary)			
6021	forward	1.41×10^{12}	18640
1,5-branching agent (primary)			
6012	forward	1.76×10^{11}	14760
1,5-branching agent (secondary)			
6022	forward	1.76×10^{11}	14760
1,5-branching agent (primary)			
6013	forward	2.20×10^{10}	12620
1,5-branching agent (secondary)			
6023	forward	2.20×10^{10}	10910

Table 2. Molecule, Research Octane Number (RON), Blend Octane Number (blend), and Initial Rate of Formation of the Branching Agent

molecule	RON	blend	rate ($\times 10^4$)
butane	94	113	0.07
pentane	62	62	1.13
hexane	25	19	2.05
heptane	0	0	2.34
2-methylpropane	100	122	0.00
2-methylbutane	92	99	0.04
2-methylpentane	73	82	0.12
2-methylhexane	42	40	0.93
2-methylheptane	20	12	1.65
3-methylheptane	27	30	1.60
4-methylheptane	27	30	1.17
3-methylhexane	52	56	1.09
2,2-dimethylbutane	92	89	0.04
2,2-dimethylpentane	93	89	0.06
2,2-dimethylhexane	73	67	0.78
2,4-dimethylhexane	65	64	0.79
2,3-dimethylhexane	71	70	0.15
3,3-dimethylhexane	76	72	0.95
3,4-dimethylhexane	76	67	0.31
3,3-dimethylpentane	81	83	0.87
2,4-dimethylpentane	83	76	0.04

counter is incremented. In addition, a filtering mechanism can be employed to eliminate unwanted cases. The algorithm is detailed in Figure 8.

4. AN EXPERIMENT USING REACTION

Reaction is an X-window based system. The reaction generation portion is part of a larger system for the manipulation and analysis of a database of molecules,

ForR1000	6.022E+10	2.610E+03	0.000E+00
ForR2010	5.420E+09	0.000E+00	0.000E+00
RevR2010	7.600E+16	1.636E+04	0.000E+00
ForR2020	5.420E+09	0.000E+00	0.000E+00
RevR2020	6.000E+15	1.708E+04	0.000E+00
ForR2030	5.420E+09	0.000E+00	0.000E+00
RevR2030	2.800E+15	1.563E+04	0.000E+00
ForR3001	1.410E+12	1.864E+04	0.000E+00
RevR3001	8.630E+10	1.200E+04	0.000E+00
ForR3002	1.760E+11	1.476E+04	0.000E+00
RevR3002	3.330E+09	8.000E+03	0.000E+00
ForR3003	2.200E+10	1.262E+04	0.000E+00
RevR3003	6.720E+09	6.170E+03	0.000E+00
ForR6011	1.410E+12	1.641E+04	0.000E+00
ForR6021	1.410E+12	1.864E+04	0.000E+00
ForR6012	1.760E+11	1.476E+04	0.000E+00
ForR6022	1.760E+11	1.307E+04	0.000E+00
ForR6013	2.200E+10	1.262E+04	0.000E+00
ForR6023	2.200E+10	1.091E+04	0.000E+00
END			

Figure 10. This is the first part of the output of the reaction generation process. It is the set of reaction constants (from the reaction database). The columns are as follows: reaction name, Arrhenius constant, activation energy, and temperature coefficient.

reactions, and reaction patterns. The system itself has a database of reactions and molecules available. To generate a reaction mechanism, one needs to specify the reactions and molecules from this database. The database is initially loaded through the use of molfiles[16] generated by hand or through ChemBase.[17] The molecules and reactions are referred to either by name or by ID number.

To generate a mechanism, three lists are specified:

Initial Molecules The initial set of molecules are specified simply by giving their database name or id.

Reaction Sequence The reaction pattern sequence is specified in a file, also by name or id, with additional information about molecules which should be present (beside those of the generated from the previous step). An example is given in Figure 9.

Flagged Molecules These are molecules, given by name or id, which are to be watched when doing the numerial calculations.

An important use of the mechanism generation process is to study the behavior of a set of molecules under similar reacting conditions. As an illustration, the simple low temperature model of combustion will be calculated for a set of hydrocarbons.

4.1. Low Temperature Oxidation Model. The low temperature oxidation reaction patterns used in this experiment are those leading to the branching agents (as initially outlined by Morley[12]). Each reaction specification can represent both the forward and reverse reactions with the corresponding kinetic data (Arrhenius constant, temperature coefficient, and activation energy).

The process toward branching agent formation, as modeled here, is represented as the following sequence of reaction-classes (an example is shown in Figure 2):

```
 1 ForR1000   M0 + M1  ->  M2 + M3      1 ForR2010   M14 + M1  ->  M24
 2 ForR1000   M0 + M1  ->  M3 + M4      4 RevR2010   M24  ->  M14 + M1
 6 ForR1000   M0 + M1  ->  M3 + M5      1 ForR2010   M16 + M1  ->  M25
 3 ForR1000   M0 + M1  ->  M3 + M6      4 RevR2010   M25  ->  M16 + M1
 1 ForR2010   M5 + M1  ->  M7           1 ForR2010   M18 + M1  ->  M26
 4 RevR2010   M7  ->  M5 + M1           4 RevR2010   M26  ->  M18 + M1
 1 ForR2010   M6 + M1  ->  M8           1 ForR2010   M20 + M1  ->  M27
 4 RevR2010   M8  ->  M6 + M1           4 RevR2010   M27  ->  M20 + M1
 1 ForR2020   M4 + M1  ->  M9           1 ForR2010   M21 + M1  ->  M28
 4 RevR2020   M9  ->  M4 + M1           4 RevR2010   M28  ->  M21 + M1
 1 ForR2030   M2 + M1  ->  M10          1 ForR2010   M22 + M1  ->  M29
12 RevR2030   M10  ->  M2 + M1          4 RevR2010   M29  ->  M22 + M1
 1 ForR3001   M7  ->  M11               1 ForR2010   M23 + M1  ->  M30
 1 RevR3001   M11  ->  M7               4 RevR2010   M30  ->  M23 + M1
 2 ForR3001   M8  ->  M12               1 ForR2020   M12 + M1  ->  M31
 2 RevR3001   M12  ->  M8               4 RevR2020   M31  ->  M12 + M1
 1 ForR3001   M9  ->  M13               1 ForR2020   M15 + M1  ->  M32
 1 RevR3001   M13  ->  M9               4 RevR2020   M32  ->  M15 + M1
 3 ForR3001   M9  ->  M14               1 ForR2020   M17 + M1  ->  M33
 3 RevR3001   M14  ->  M9               4 RevR2020   M33  ->  M17 + M1
 2 ForR3001   M10  ->  M15              1 ForR2030   M11 + M1  ->  M34
 2 RevR3001   M15  ->  M10             12 RevR2030   M34  ->  M11 + M1
 6 ForR3001   M10  ->  M16              1 ForR2030   M13 + M1  ->  M35
 6 RevR3001   M16  ->  M10             12 RevR2030   M35  ->  M13 + M1
 2 ForR3002   M7  ->  M17               1 ForR2030   M19 + M1  ->  M36
 2 RevR3002   M17  ->  M7              12 RevR2030   M36  ->  M19 + M1
 3 ForR3002   M7  ->  M18               1 ForR6011   M31  ->  M37 + M38
 3 RevR3002   M18  ->  M7               1 ForR6011   M34  ->  M38 + M39
 1 ForR3002   M8  ->  M19               1 ForR6021   M24  ->  M38 + M40
 1 RevR3002   M19  ->  M8               1 ForR6021   M35  ->  M38 + M41
 6 ForR3002   M9  ->  M20               1 ForR6012   M26  ->  M38 + M42
 6 RevR3002   M20  ->  M9               1 ForR6012   M33  ->  M38 + M43
 3 ForR3002   M10  ->  M21              1 ForR6012   M36  ->  M38 + M44
 3 RevR3002   M21  ->  M10              1 ForR6022   M27  ->  M38 + M45
 3 ForR3003   M7  ->  M22               1 ForR6013   M29  ->  M38 + M46
 3 RevR3003   M22  ->  M7               1 ForR6013   M30  ->  M38 + M47
 6 ForR3003   M8  ->  M23
   RevR3003   M23  ->  M8
```

Figure 11. The second part of the output of the reaction generation process is the mechanism (2-methylbutane is shown here). The multiplicity of the reaction, the reaction name (refering to the reaction constants), and the reaction are specified.

Initiation   The initial reaction is hydrogen abstraction by oxygen producing a peroxide and an alkyl radical (reaction 1000). Dependence on primary, secondary, and tertiary carbons is left to the next step. Only the forward direction is considered.

Alkylperoxy Formation   The addition of oxygen to the radical to give an alkyperoxy radical is one of the first propagating reactions. The rate constants used are dependent on carbon type[15] (reactions 2010, 2020, and 2030 for primary, secondary, and tertiary carbon radicals). Both the forward and reverse directions are condisered.

Isomerization   A set of (1,4), (1,5), and (1,6) isomerisations leading to alkylperoxides is the next step toward the production of the branching agent. The constants for primary hydrogen abstraction are used[15] (reactions 3001, 3002, and 3003). Both the forward and reverse directions are considered.

Second O₂ Formation   The addition of oxygen here is treated as the first alkylperoxy formation, i.e., no dependence on the presence of the peroxy group.

Aldehydes   The only type of branching agent formation considered: A concerted reaction of hydrogen abstraction and formation of two hydroxyl radicals. The rate constants used are those of the forward direction of the isomerisation step (reactions 6011, 6021, 6012, 6022, 6013, and 6023).

One of the main purposes of this experiment is the study of the production of branching agents. The reactions were chosen such that the production of hydroxyl radicals is associated with the production of branching agents. It is the set of the hydroxyl radicals which will be the indicator of the rate of branching agent formation. Associated with each reaction is a set of rate constants. These are outlined in Table 1.

These five steps are performed sequentially. The initial hydrocarbon undergoes the initiation (nonreversible) step of
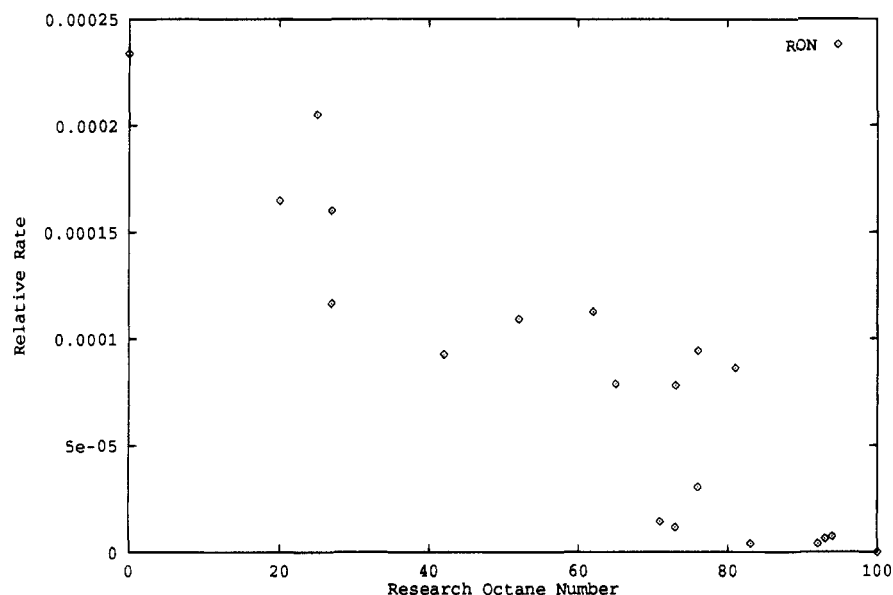
Figure 12. The calculated branching agent rate versus research octane number is plotted for the series of hydrocarbons shown in Table 2.
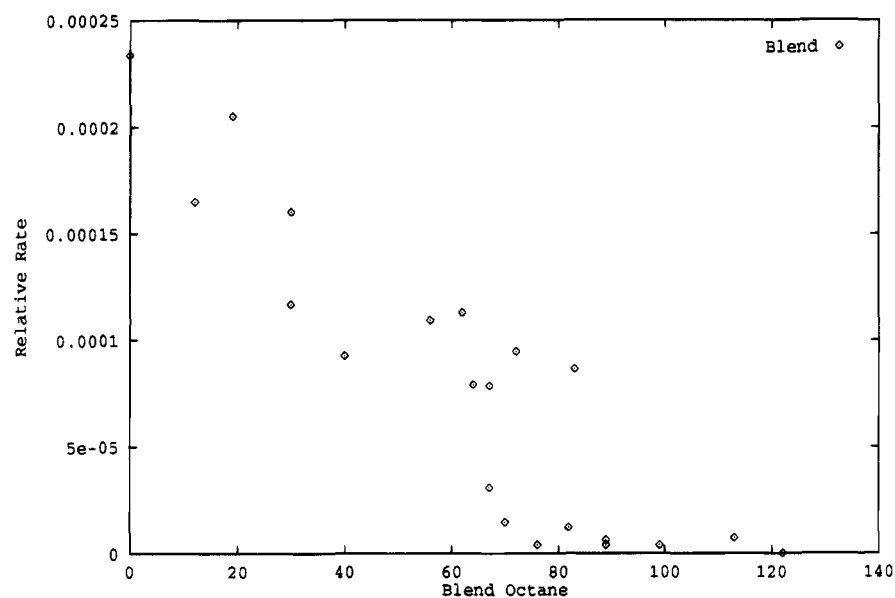

Figure 13. The calculated branching agent rate versus blend octane number is plotted for the series of hydrocarbons shown in Table 2.

hydrogen abstraction. The resulting intermediate hydrocarbon radicals then undergo the second step of addition of oxygen. The original hydrocarbon (and the oxygen) is not sent to the next step. Only newly generated intermediates are allowed to react further. The next reaction set is three reaction patterns for primary, secondary, and tertiary carbon addition of oxygen. These are applied within the current set of intermediates. Oxygen must be given as an additional molecule (see Figure 9), because it is not generated by the previous step. The resulting hydrogen peroxyl radicals are then sent to the next set of reactions in the sequence. At the end of this sequence of reaction patterns is a set branching agents and a hydroxyl radical. Two hydroxyl radicals are formed for each branching agent.

4.2. The Generated Reaction Mechanism. The form of the output of the reaction mechanism is such that further analysis can be performed with a numerical tool for kinetic analysis. The output basically consists of three parts:

Reaction Constants The reaction constants (taken from the original database) for the set of reactions are given in a list. The constants are the Arrhenius constant, the activa-

tion energy and the temperature coefficient. Figure 10 gives the set used here.

Mechanism The mechanism is given as a list of individual reactions with the multiplicity, the reaction name, and the molecules involved. Figure 11 gives the reaction mechanism for 2-methylbutane.

Flagged Molecules Since the molecules are named generically with M*x*, where *x* is an integer, a correspondence to the set of molecules that the user would like to watch is also given.

4.3. Using the Mechanism in Kinetic Analysis. The reaction mechanism is applied to a set of molecules. The rate of formation of the branching agent in this experiment is followed by the formation of the hydroxyl radical. In *Reaction*, the entire process runs automatically from generation of the mechanism to the generation of the output graphs:

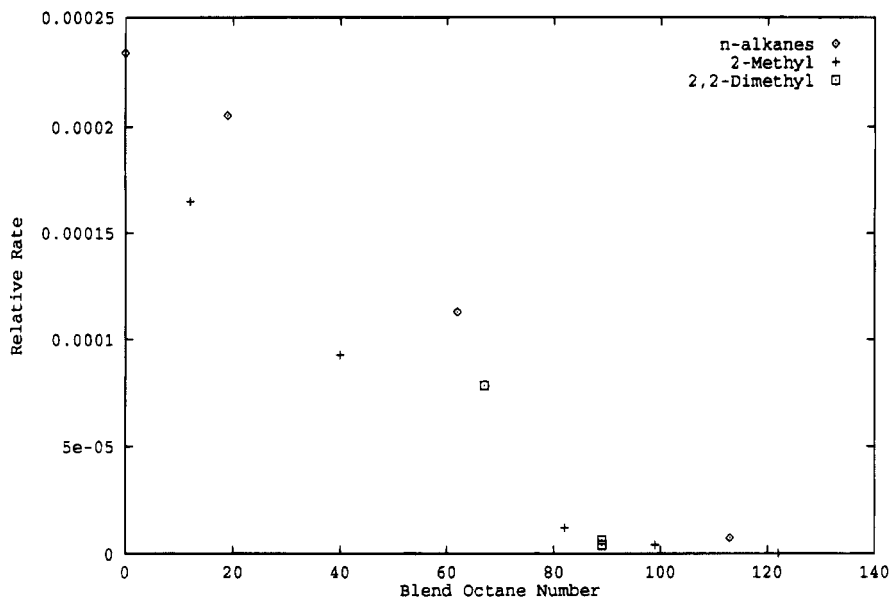**Mechanism** The mechanism is generated and given in the form shown in Figures 10 and 11.

Figure 14. Blend octane plot for *n*-alkanes, 2-methylalkanes, and 2,2-dimethylalkanes.
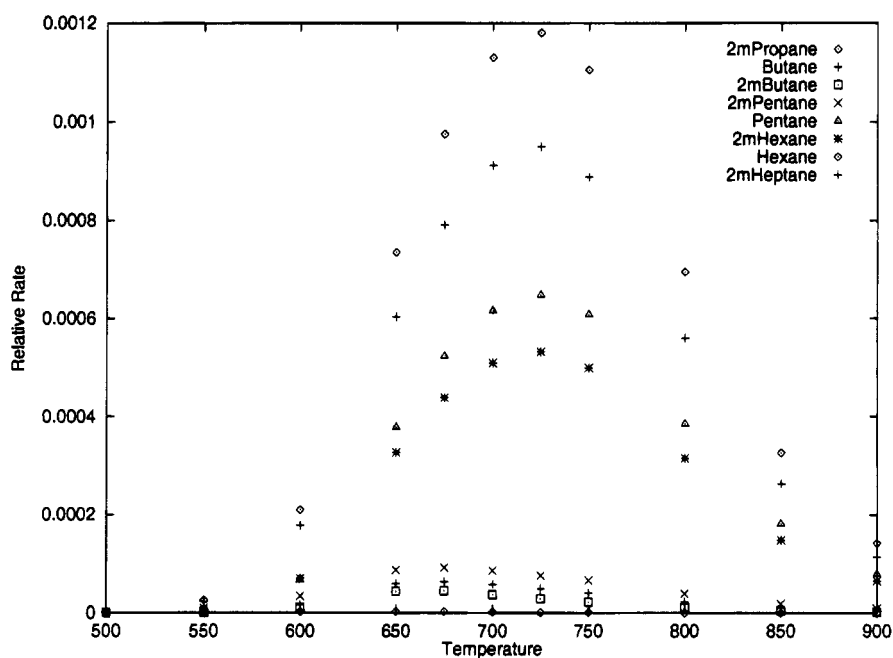

Figure 15. A plot of the rate of hydroxyl (branching agent) production versus temperature for the following hydrocarbons (with RON in parentheses): 2-methylpropane (100), butane (94), 2-methylbutane (92), 2-methylpentane (73), pentane (62), 2-methylhexane (42), hexane (25), and 2-methylheptane (20).

Setup	The data of the mechanism is converted to a form suitable for input to the kinetic analysis routines, REACT, of Whitbeck.[18]
Kinetics	REACT is used to generate the time dependent concentration behavior of the hydroxyl radical.
Summary	As soon as all the mechanisms have been evaluated, the data for each are collected and summarized in table form (with the information shown in Table 2) and plotted with gnuplot.

The purpose of the experiment is to reproduce the somewhat linear trend, observed by Morley[12], between the production rate of branching agents and the blend octane number. As Table 2 and Figures 12 and 13 show, a vague linear trend, enough for qualitative substantiation of the method, but not enough for quantitative utility is found.

Figure 14 shows that the trend for a simple series of hydrocarbons, namely the *n*-alkanes, does show a very linear relationship. The 2-methylalkane trend is not linear but apparently smooth. However, it should be mentioned that such qualitative trends are quite easy to come by when dealing with such restricted classes of molecules. For example, for the *n*-alkanes, the increase in chain length can be represented in a variety of ways, from counting hydrogens and/or carbons to more complex analyses as this one. It can be seen more as a combinatorial effect of the consistent increase of the same structure group, CH_2.

In this model, there seems to be a consistent underestimation of the effect of substitution. The first hint of this is the trend of the larger 2-methylalkanes which is somewhat linear but consistently lower than that for the unsubstituted *n*-alkanes. The somewhat flat relationship between the octane numbers and the rates of the dimethyl substituted alkanes

616 *J. Chem. Inf. Comput. Sci., Vol. 35, No. 3, 1995*

BLUROCK

and their consistently low values substantiates this point. From these simple observations one could conclude either that *n*-alkane rates are overestimated or the substituted alkanes are underestimated. Assuming the latter, this would probably mean that an important channel toward branching agent formation has not been considered for branched alkanes (assuming also, of course, that branching agent formation is still the significant step). Further experimentation with additional schemes will be studied in the future.

Figure 15 is a plot of the initial rate of branching agent formation versus temperature for a set of 2-methylalkanes. The "bell-shaped" curve indicates the typical behavior of this simplified mechanism. The key to this behavior is the reversible addition of oxygen to the radical hydrocarbon. At higher temperatures, the reverse reaction becomes significant and "shuts off" the production of branching agents.

As one last note, one cannot neglect the fact that the kinetic study done here is overly simplistic not only in the choice of mechanism but also in the other physical factors that definitely play a role; the most important of which would be heat evolution. Overcoming both of these shortcomings is planned for the future. This work represents the very beginnings of a system to automatically generate and evaluate mechanistic schemes for a wide class of molecules.

5. CONCLUSION AND FUTURE DIRECTIONS

Reaction is at the beginning of its development. But even at this early stage it can yield useful information about reaction mechanisms. Its main advantage is that the user can characterise the entire detailed mechanism in terms of a small set of reaction prototypes. The details of expanding these to the full set of specific reactions for a specific molecule is done automatically. The set of resulting reactions (and their corresponding reaction constants) are then available for use in, for example, further kinetic analysis.

The author sees the advantage of such a system in the rapid generation of a consistent (i.e., under the same set of reaction types) set of reactions for a set of molecules for comparison. Further advantages will arise when further information within the reaction tree is also automatically generated, for example, kinetic and thermodynamic constants (entropy, enthalpy and heat capacities). Simple schemes already exist which, given an arbitrary molecule or radical, generate these constants.[19-22] In future work these schemes will be bound to the system so that the system will not only generate the reaction sets but also deliver the necessary thermodynamic and kinetic data for complete kinetic analysis (for example, with the CHEMKIN system[23]). The most significant advantage, in the view of the author, will be the use of reaction patterns in the study and generation of automatic lumping schemes.

REFERENCES AND NOTES

(1) Wilcox, C. S.; Levinson, R. A. A Self-Organized Knowledge Base for Recall, Design and Discovery in Organic Chemistry. *ACS Symp. Ser.* **1986**, *306*, 209−230.

(2) Blurock, E. S. Computer-aided Synthesis Design at RISC-Linz: Automatic Extraction and use of Reaction Classes. *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 505−510.

(3) Gasteiger, J.; Hutchings, M. G.; Christoph, B.; Gann, L.; Hiller, C.; Loew, P.; Marsili, M.; Saller, H.; Yuki, K. A New Treatment of Chemical Reactivity: Development of EROS, an Expert System for Reaction Prediction and Synthesis Design. *Top. Curr. Chem.* **1987**, *137*, 19−72.

(4) Bawden, D. Classification of Chemical Reactions: Potential, Possibilities, and Continuing Relevance. *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 212−216.

(5) Funatsu, K.; Endo, T.; Kotera, N.; Sasaki, S. Automatic Recognition of Reaction Site in Organic Chemical Reactions. *Tetrahedron Computer Methodology* **1988**, *1*, 53−70.

(6) Dente, M.; Pierucci, S.; Ranzi, E.; Bussani, G. New Improvements in Modeling Kinetic Schemes for Hydrocarbons Pyrolysis Reactors. *Chem. Eng. Sci.* **1992**, *47*, 2629−2634.

(7) Ranzi, E.; Faravelli, T.; Gaffuri, P.; Sogaro, A. Low Temperature Combustion: Automatic Generation of Primary Oxidation Reaction and Lumping Procedures. *25th International Symposium on Combustion*; Irvine, CA, 1994.

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

(9) Chevalier, C.; Pitz, W. J.; Warnatz, C. K.; Westbrook, C. K.; Melenk, H. Hydrocarbon ignition: Automatic generation of reaction mechanisms and applications to modeling of engine knock. *24rd Symposium (International) on Combustion Proceedings*; Combustion Institute: Pittsburgh, PA, 1992; pp 93−101.

(10) Chinnick, S. J.; Baulch, D. L.; Ayscough, P. B. An expert system for hydrocarbon pyrolysis. *Chemometrics Intelligent Laboratory Systems* **1988**, *5*, 39−52.

(11) Cox, R. A.; Cole, J. A. Chemical Aspects of the Autoignition of Hydrocarbon−Air Mixture. *Combustion Flame* **1985**, *60*, 109−123.

(12) Morley, C. A Fundamentally Based Correlation Betwen Alkane Structure and Octane Number. *Combust. Sci. Tech.* **1987**, 115−123.

(13) Frenklach, M. Reduction of Chemical Reaction Models. In *Progress in Astronautics and Aeronautics*; Oran, E. S., Boris, J. P., Ed.; 1988; Vol. 135; pp 129−151.

(14) Cowart, J. S.; Keck, J. C.; Heywood, J. B.; Westbrook, C. K.; Pitz, W. J. Engine Knock Predictions Using a Fully-Detailed and a Reduced Chemical Kinetic Mechanism. In *23rd Symposium (International) on Combustion Proceedings*; Combustion Institute: Pittsburgh, PA, 1990; pp 1055−1062.

(15) Griffiths, J. F.; Hughes, K. J. A Unified Approach to the Reduced Kinetic Modelling of Alkane Combustion. In *Combustion Flame* **1994**, *99*, 533−540.

(16) *REACCS: Reference Manual*, Molecular Design, Ltd., San Leandro, CA, U.S.A.

(17) *ChemBase: Chemist's Personal Software Series*, Molecular Design, Ltd., San Leandro, CA, U.S.A.

(18) Whitbeck, M. REACT−Program to solve kinetic equations for chemical systems (version 1.00). Technical report, Desert Research Institute, 1991.

(19) Benson, S. W. *Themochemical Kinetics*, 2nd edition; John Wiley and Sons: New York, 1976.

(20) Downey, J. R.; Frurip, D. J.; LaBarge, M. S.; Syverud, A. N.; Grant, N. K.; Marks, M. D.; Harrison, B. K.; Seaton, W. H.; Treweek, D. N.; Selover, T. B. The ASTM Computer Program for Chemical Thermodynamic and Energy Release Evaluation: Chetah 7.0 Users' Guide; Technical Report, National Institute of Standards and Technology, Gaithersburg, MD, 1994.

(21) Dean, A. M.; Bozzelli, J. W. CHEMACT. *Combust. Sci. Tech.* **1991**, *80*, 63−85.

(22) Bozzelli, J. W.; Ritter, E. R. RADICALC: A Computer Code to Estimate Entropies and Heat Capacities for Radical Species and Transition State Structures. In preparation.

(23) Kee, R. J.; Miller, J. A.; Jefferson, T. H. *CHEMKIN: A General-Purpose, Problem-Independent, Transportable, FORTRAN Chemical Kinetics Code Package*. Sandia Laboratories Report, SAND80-8003 1980.