Chapter 3
Automatic Generation of Detailed Mechanisms

**Edward Blurock, Frédérique Battin-Leclerc, Tiziano Faravelli
and William H. Green**

Abstract The production of detailed mechanisms for the combustion of larger hydrocarbons is fundamentally different from the production of mechanisms for the reactions of smaller hydrocarbons. For the most part there are no direct experiments for the individual reactions. General physical principles, usually based on reaction classes, have to be employed. In addition, the production of larger mechanisms has to be careful and systematic to generate what could be mechanisms with thousands of species and reactions. The recurring and systematic use of physical principles to generate the mechanisms is facilitated by automated techniques in the form of automatic detailed mechanism generation systems. This chapter outlines the common principles, with some software technical details, of automatic generation systems. This is followed by a more detailed description of four individual systems, MAMOX, EXGAS, RMG and REACTION, each of which has unique aspects in their generation philosophy.

E. Blurock
Reaction, Lund, Sweden
e-mail: edward.blurock@esblurock.info

F. Battin-Leclerc (✉)
Laboratoire Réactions et Génie des Procédés, CNRS, Université de Lorraine, ENSIC, 1 rue grandville, Nancy 54000, France
e-mail: frederique.battin-leclerc@univ-lorraine.fr

T. Faravelli
Dipartimento di Chimica, Materiali e Ingegneria Chimica, Politecnico di Milano, Piazza L. da Vinci, 32, Milan 20133, Italy
e-mail: tiziano.faravelli@polimi.it

W. H. Green
Department of Chemical Engineering, Massachusetts Institute of Technology 66-350, 77 Massachusetts Avenue, Cambridge, USA
e-mail: whgreen@mit.edu

3.1 Common Principles of Automatic Generators

As detailed mechanisms of larger hydrocarbons grow larger, so does the complexity of their production. Although it is possible to produce large mechanisms of thousands of reactions by hand (see, for example, the work of (Westbrook et al. 2009; Sarathy et al. 2011)), having an automatic means of generation can be more efficient, more systematic and less error-prone. Automatic detailed mechanism generators can be viewed as expert systems using a database of chemical principles to systematically and efficiently produce large detailed mechanisms. One of the principle advantages of automatic generators is that the time-consuming and error-prone details of producing every single species and reaction in a large mechanism is taken over by the generation system. The modeler works at a higher conceptual level determining, for example, which submechanisms and which *reaction classes* should be generated. The generation procedure itself is more systematic which is important not only because of reduction of errors, but also in the consistency of the use of general kinetic principles, usually in the form of reaction classes, that are applied.

3.1.1 Core Structure of an Automatic Generator

Regardless of the particular implementation of an automatic generator, there are distinct commonalities in their structure. Within this common structure, there are several design and strategic decisions where each system can differentiate. Every generator has a set of core modules consisting of the *generator engine*, the *species pool*, the *molecule database* and the *reaction class database*:

- *Generator engine*. The central module, interacting with all the other modules, which steers the generation process.
- *Species pool*. These are the molecules that, under each iteration, serve as input to the generator engine to produce the reactions and molecules (taken in a wide sense including both stable molecules and free-radicals) of the current iteration.
- *Molecule database*. This is the set of predefined molecules that could be used within the generation process.
- *Reaction class database*. This database contains the information about each reaction class to be used in the generation process.

The iterative use of these modules by the reaction generation algorithm creates the reactions and species making up the generated mechanism. The input and output to the algorithm is:

- *Input*. The fuel molecule or set of fuel molecules.
- *Output*. The detailed mechanism associated with the fuel molecule(s).
  The algorithm itself is iterative and has the following basic steps:
- *Initialization*. Fill the reaction pool with initial reactants.

- *Generate reactions.* Using the information in the reaction class database, create a new set of reactions and molecules using the species pool as input. Add the generated molecules and reactions to the mechanism.
- *Update the species pool.* Update the species pool with the newly generated species from step 2.
- *Check termination criteria.* If the algorithm is done, then exit, otherwise iterate to step 2.

The following sections elaborate these steps and outline the different strategies each step can have.

3.1.1.1 Initialization

The input to the algorithm is the fuel molecule or molecules. In the initialization step, these species are added to the species pool. The submechanism that is generated is considered as the *primary mechanism* or reaction[1] of those initial molecules. This generated submechanism is then used in conjunction with other submechanisms (which can be generated or produced by hand). The *Generate Reactions* step uses the species pool as input to generate the next set of reactions. It could be that in the initialization process, some "extra" reactants, such as small radicals used in some reactions have to be added to the species pool. These are taken from the molecule database.

3.1.1.2 Generate Reactions

In this step, the species pool is analyzed and those reaction classes from the reaction class database which are valid are applied. A reaction class is valid for a species (or set of species) if the *functional groups* required by the reaction class are present. The recognition of the necessary functional groups is done through some form of graph isomorphism (see Sect. 3.1.3.4 and the general references (Balaban 1985, 1995)). The exact algorithm is highly dependent on the molecular representation (see Sect. 3.1.3). There could be some filtering done at this stage as to which reaction classes are available for application.

Each application of the reaction class produces a set of reactions and a set of molecules. This new set of reactions and molecules are collected. A crucial algorithm in this step is the determination of which molecules and reactions are equivalent (see Sect. 3.1.3). Only newly generated reactions and molecules are added to the new mechanism. During this generation step, there could be some filtering of applications and results. The filter could occur before or after the

---

[1] The *primary mechanism* includes the reactions of the reactants and of the directly derived radicals.

reaction class application. There could be other conditions, beside the availability of the functional groups which inhibit the application of the reaction class. For example, if the reaction class is known to generate a radical center and the reacting species is already a radical, the generation of a biradical could be inhibited. The filtering could also occur after the application of the reaction class. For example, the generated reaction and associated molecules may have some criteria they must fulfill. One class of such criteria could be based on "on-the-fly" kinetic or thermodynamic computations (this, for example, is crucial in the RMG system described in Sect. 3.2.3). The reactions are left in the final set only if they meet these criteria.

3.1.1.3 Update the Species Pool

The species pool is updated with the new set of generated species. There could be some filtering and even classification of the incoming species. There are essentially two algorithms to update the species pool:

- *Exhaustive*. The (filtered) set of generated molecules is added to the species pool, creating a larger pool.
- *Progressive*. The species pool is initialized and the (filtered) set of species is added at each step.

Under the exhaustive technique, the reactions are applied (in the Generate reaction step) until no new species are produced. This check is made after filtering. In the progressive technique, only selected species are used in the next step. For example, in the REACTION system (see Sect. 3.2.4) only the newly generated species are added at each step of the defined reaction pathway.

3.1.1.4 Termination Criteria Check

If no new species were added to the species pool then the process can terminate. This essentially means that all of the available or required reaction classes have been applied and no new reactions and molecules were created. This could result from the fact that new applications of the reaction classes give molecules and reactions already in the mechanism, or that there are no more reaction classes to apply.

3.1.2 Convergence and Termination of the Generation Process

As discussed in Chap. 2, a typical combustion mechanism is a hierarchy of sub-mechanisms. The "top" mechanism, for example that of the original fuel, starts

with a large molecule and in the reactive process produces smaller, usually containing less carbon atoms, species. These smaller species are consumed in subsequent sub-mechanisms of the hierarchy. In general, when combustion mechanisms are designed and the individual reactions chosen, regardless of whether they are automatically generated or not, for the most part the species are reacted into smaller species. A particular submechanism is designed so that its products are consumed by other sub-mechanisms.

One of the challenges of *low temperature oxidation* of hydrocarbons (see Chap. 2) was that an addition, for example the addition of oxygen to an alkyl radical had to be modeled. However, this addition is allowed because subsequent reaction classes once again reduce or keep constant the size of the products, at least in terms of number of carbon atoms. Fortunately, the product species obtained, for example alkylperoxy radicals, have very specific structures (functional groups) so they can be targeted by very specific reaction classes which will eventually react them to smaller species. Recombinations also increase the size of the species in the generation process. If additions of hydrocarbon radicals or recombinations are allowed in the (automatic) generation process larger and larger molecules would be available to react further. The process would then only terminate if a maximum species size is specified, this is the case in EXGAS software (see Sect. 3.2.2.4). This problem would be even more severe for soot formation, where molecular growth is very important.

A terminating iterative process converges to the solution after a finite number of steps. One of the basic criteria can be that each step creates a "state" that is smaller than the original state. In mechanism generation the current state is the molecule to be reacted. One possible definition of the size of the state is the size of the species to react. After the application of a reaction class, smaller species are generated and the process is one step closer to convergence to the final mechanism. If a reaction class produces a species of the same size, then the generation process has to have some means of limiting the applications of that class to a finite number of times. Without this check, a reaction class such as isomerization would produce an infinite loop. One check to this process is to see whether the same molecule has already been produced. Since a given molecule can only have a finite number of isomers, a reaction class producing another isomer can only be applied a finite number of times (though the number of isomers can be very large for large molecules).

3.1.3 Molecular and Reaction Representations

As said in Chap. 2, a combustion modeler uses many representations of molecule and subsequently reactions. These models help the human modeler understand the reactive process of combustion. To be useful in automatic generation these models have to be translated into machine readable data structures. These data structures go beyond the classical numerical representations that are normally associated with a computer aided computation. The evolutionary development of these

structures coincides with developments of modern computing techniques and languages (Chen 2006; Willett 2008). In addition, the representations, structures and many of the techniques used within automatic generation of combustion mechanisms stem from the extensive work in physical and organic chemistry in general, especially in the field of chemical information databases.

3.1.3.1 Molecule Names for Humans and Computers

A molecule is a complex object with many models and representations. A fundamental task in organic and subsequently combustion chemistry is to give each molecule a unique name. Two molecules with the same unique name can be said to be identical. In fact, assigning unique names or codes to a graph is one of the more computationally efficient ways to identify identical graphs. Unfortunately, the task of naming is still complex.

Naming a molecule, for the most part, uses the Lewis structure (see Chap. 2 and, for example (Gillespie and Robinson 2007; Lewis 1916)) and a usually used representation of the molecule is a two-dimensional (2D) graph (Balaban 1985, 1995). The atoms are graph nodes with additional information such as atomic number, charge, radical, etc. The graph bonds are essentially the full covalent bonds of the molecule. In general, finding a unique name, or more precisely a canonical form, for a graph is a complex task (Babai and Luks 1983) and is intimately connected to identifying if graphs are identical or not (graph isomorphism (Balaban 1985)). Within the field of chemistry these tasks are extremely important for molecular databases and data-mining chemical information. Fortunately, organic compounds, using the Lewis model, are usually not as complex as a general graph that has to be dealt with mathematically. For example, sp^3 atoms have at most four bonds which greatly simplifies the interconnectivity among graph nodes.

In organic chemistry standardization of names is an important task and the International Union of Pure and Applied Chemistry (IUPAC) has undertaken the standardizing of nomenclature. Having a standardized (canonical) name not only tells the organic chemist what the molecule looks like, but also simplifies searching. Two molecules are alike if the text string of their name is exactly the same. For example, the species *iso*-octane has the IUPAC name of 2,2,4-trimethylpentane. The IUPAC name is an example of linearizing the graphical structure of the molecule. Another, more compact linear name for molecules are the Wiswesser Line Notation (WLN) (Smith et al. 1968; Wiswesser 1982, 1985). Several variants such as the Simplified Molecular Input Line Entry System (*SMILES*) notation (Weininger 1988; Weininger et al. 1989), the SYBL line notation (Homer et al. 2008) and the *linear notation* (Côme et al. 1984; Warth et al. 2000 (see Sect. 3.2.2.1)) have also been proposed. These notations were mainly developed for simplified ASCII text entry of graphical molecular information into the computer. These notations represent the molecule as a "spanning tree" with the use of parenthesis to denote branching. Some notations rely on

molecular valence information to simplify the notation. For example, in the SMILES notation, hydrogen atoms are implicit.

Another human readable name for a molecule is the International Chemical Identifier (InChI) (Stein et al. 2003), "InChI" for short, which was developed by IUPAC and National Institute of Standards to facilitate molecular searches in databases and on the web. The current nomenclature has 6 layers of structural information about molecule. This is becoming a standard nomenclature for molecular searches on the internet.

3.1.3.2 Bonding Representations

The representation of molecular species for human–computer interfacing and for internal representation is not unique to combustion (Warr 2011) and is constantly evolving with the needs of the chemistry community in chemical database search, computer aided organic synthesis (Corey and Wipke 1969; Wipke and Howe 1977) and chemoinformatics (Engel 2006; Willett 2008). Of course, some of these needs have evolved hand in hand with developing software technologies (Engel and Gasteiger 2002; Chen 2006).

The naming of the chemical species, whether it be with the common name, IUPAC name or one of the linear notations, is an efficient way to get molecular data into the automatic generation system. The main purpose of these notations is that the name gives a direct correspondence to the molecular structure needed. However, within the computations needed for automatic generation these forms are not efficient. Though, for example, a canonical name is efficient for recognizing whether two species are equal. They are less efficient at recognizing, for example, whether a given substructure is within another species structure or the transformations of species creating a reaction.

As with the naming of species, the essential information in the main internal data structures for molecules consists of the Lewis model information, meaning a 2D graph representation. If additional molecular data is needed, it is usually added in a separate data structure. In computer science there has been essentially two ways to represent a graph both internally and for ASCII human–computer interfacing: a connection table (or matrix) or a graph as a set of atoms and a set of bonds between the atoms.

Historically, one can say one that the first computational representation of molecules was the atom connectivity matrix (sometimes also called the adjacency matrix) (Spialter 1963, 1964). This form was conducive to numerical programming languages such as FORTRAN. A species with n atoms would be represented as a matrix with n rows and n columns. A connection is signified by a non-zero value in the off-diagonal elements. Ugi applied this to molecules with the definition of the Bond-Electron (BE) matrix (Dugundji and Ugi 1973; Ugi and Wochner 1988). If a bond exists between atom i and atom $j,$ then element m_{ij} of connectivity matrix $\mathbf{m}$ has the order of the bond (1 for single, 2 for double, 3 for triple). In the era where computation was primarily numerical, an interesting

extension of the BE-matrix was the R-matrix, the difference between the product and reactant BE-matrix. The significance of the R-matrix is that it is an early representation of reactive changes that could be used directly to "calculate" product species from the reactant species which is an essential operation in automatic generation.

The connection table is a form which is close to the 2D graphical form and, for example, can be used as the internal representation and as ASCII input. The table usually consists of two parts, atom information and bonding. The exact syntax can vary, first due to historical developments and second due to individual needs of the software systems using it. Some of the more accepted variants have been developed by Molecular Design Limited (MDL) (Dalby et al. 1992). The description of the molecule has a fixed format (originally designed for FORTRAN-like formating) and has at least two parts, the first being the atomic description and the second the bonding description. Addition parts can include additional information. This representation can basically be translated one-to-one to an internal 2D-graphical data structure.

A more general and modern relative to the MDL connection tables is the Chemical Markup Language (CML) (Murray-Rust and Rzepa 1999) which is a specialization for physical and organic chemistry of the Extensible Markup Language (XML). XML is a standard set up by the World Wide Web Consortium (W3C) for the transmission of data over the internet that is both human and machine readable. The advantage of a format based on XML is that a wide range of software in a wide range of languages has been independently developed. This includes a set of software (written in JAVA) that has been developed specifically for chemical applications, namely the Chemical Development Kit (CDK). There is a chemical data structure directly associated with the CML format. Some automatic generators are already using these standard softwares.

3.1.3.3 Canonicity and Molecular Equivalence

An important operation in automatic generation is to determine whether two molecules are the same or not. For example, in one of the recursive steps when a molecule is generated, it is important to recognize whether the molecule already exists in the species pool. During the generation process, molecular equivalence is used to determine whether two reactions are the same. Furthermore, since a generated mechanism is usually combined with other mechanisms, for example the base mechanism or another generated mechanism, it is important to identify equivalence.

A simple way to detect whether two molecules are equivalent is if they have equivalent names. However, looking at the complexity of the IUPAC nomenclature rules, determining the canonical (unique) name is non-trivial. For example, from the names 2,2,4-trimethylpentane and 4,2,2-trimethylpentane, a modeler could draw the same structure. However, only the first is the IUPAC name and one can not use the two names to textually see whether the two structures are the same. Even a simple molecule such as *n*-pentane can be written in four different ways in

the SMILES representation: CCCCC, C(C)CCC, C(CC)CC, C(CCC)C. Though if the extra computation is made to order the atoms for the SMILES notation (Morgan 1965), then the SMILES string can be used to identify equivalent molecules. A general algorithm for the identification of equivalent structures is the graph isomorphism algorithm (Balaban 1985). The full graph isomorphism algorithm yields an atom–atom correspondence between the graph molecules. If all the atoms in both molecules can be matched, then the molecules are equivalent (the algorithm can stop after the first pair is found).

Figure 3.1 shows an example with isooctane (2,2,4-trimethylpentane). Due to symmetry C_a, C_c and C_d atoms and, correspondingly, C_1, C_3 and C_4 atoms are equivalent. In addition, C_g and C_h atoms and, correspondingly, C_7 and C_8 atoms are equivalent. This results in $((3 * 2 * 1) * (2 * 1))$ or 12 unique ways to pair atoms in the two molecules up.

3.1.3.4 Substructure Search

An essential part of recognizing whether a reaction class can be applied to a molecule is to recognize whether the essential functional group of the reaction class can be found in the molecule. This operation is highly dependent on representation of the reaction class and the molecule. If these representations are based on 2D-graphs, as with most automatic generators, then the general algorithm used is once again graph isomorphism (Ullmann 1976; Barnard 1993). Within some automatic generators the complexity of the graph isomorphism algorithm is reduced by making use of additional information and specialized graphs (such as dealing with a tree instead of a graph, as it is the case of EXGAS software, see Sect. 3.2.2.1) available during the search.

Suppose we wish to perform a hydrogen atom abstraction at a primary carbon on *n*-butane. Figure 3.2 shows the structures involved, the primary carbon atom, represented as a general methyl group, and *n*-butane molecule. The result of the

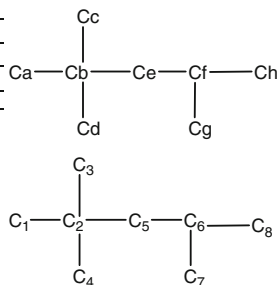|	Combinations											
	Set a1				Set a2				Set a3			
	c3		c4		c1		c4		c1		c3	
	g7	g7	g8	g8	g7	g7	g8	g8	g7	g7	g8	g8
a	1	1	1	1	3	3	3	3	4	4	4	4
b	2	2	2	2	2	2	2	2	2	2	2	2
c	3	3	4	4	1	1	4	4	1	1	3	3
d	4	4	3	3	4	4	4	4	3	3	1	1
e	5	5	5	5	5	5	5	5	5	5	5	5
f	6	6	6	6	6	6	6	6	6	6	6	6
g	7	7	7	7	7	7	8	8	7	7	8	8
h	8	8	8	8	8	8	7	7	8	8	7	7

Fig. 3.1 Atom-to-atom correspondences between two *iso*-octane molecule graphs. There are twelve combinations due to symmetry. The sets are grouped according to how Ca, Cc and Cg atoms are matched

graph isomorphism, i.e. the atom-to-atom correspondence between the graphs, is shown in the table. Due to the symmetry of the methyl group, there are (3 * 2 * 1) or 6 ways to match the methyl group at each of the two primary carbon atoms. Note first that only two sets of atoms are matched, those on carbon atom 1 and carbon atom 4. Each of the combinations involves the same atoms meaning there are two sets of three equivalent hydrogen atoms. Each of these sets has six combinations of atom-to-atom matches due to the combined symmetry of the methyl group and the primary group on the *n*-butane molecule. This redundancy, due graph symmetry, has to be taken into account to come up with the desired result of abstracting from a carbon atom from each end of the *n*-butane molecule.

To illustrate the role of reaction class structure symmetry, suppose the hydrogen abstraction is defined by removing H_b in Fig. 3.2 and the abstraction rate is defined per hydrogen atom. This means that six hydrogen atoms can be abstracted from butane. In the table these are the atom-to-atom combinations labeled: b5, b6, b7, b12, b13 and b14. Due to the symmetry of the methyl group, there are two combinations which would yield the same result. Each of these pairs would be combined to create one reaction. In total, six reactions would be generated. However, due to the symmetry of the methyl group and the symmetry of the *n*-butane molecule, all of these six reactions are equivalent. These would be recognized and then combined to one reaction with 6 times the rate of each one (the per hydrogen atom rate).

3.2 Systems Descriptions

The previous sections outlined the general principles and some algorithms of automatic mechanism generation. In this section several systems which have been applied to model combustion chemistry will be described. Each system has its own generation philosophy and database. On a software technical level, mainly due to their historical development parallel to the development of software systems in general within computer science, there are differences not only in the computer languages involved, but also the internal representations of the chemical and kinetic elements needed for generation. All systems described have fundamental
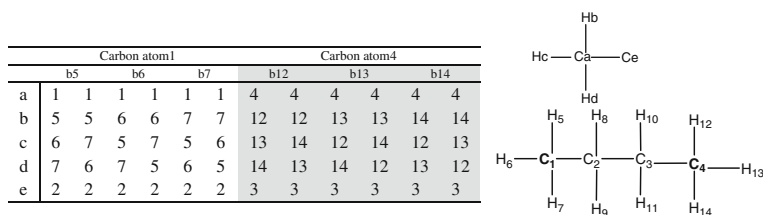
	Carbon atom1						Carbon atom4					
	b5		b6		b7		b12		b13		b14	
a	1	1	1	1	1	1	4	4	4	4	4	4
b	5	5	6	6	7	7	12	12	13	13	14	14
c	6	7	5	7	5	6	13	14	12	14	12	13
d	7	6	7	5	6	5	14	13	14	12	13	12
e	2	2	2	2	2	2	3	3	3	3	3	3

Fig. 3.2 Atom to atom correspondences for a primary methyl group in *n*-butane. There are 12 combinations

aspects which are in common. But all have particular aspects and stategies with particular advantages. In this section four automatic generation systems will be discussed:

- *MAMOX++* (Ranzi et al. 1995a). This system distinguishes itself by producing a hierarchy of (highly) lumped mechanisms derived numerically from automatically generated detailed mechanisms. The *MAMOX++* program is derived from the MAMA program in the SPYRO system (Dente et al. 1979, 1992) which was the first to automatically generate the pyrolysis mechanisms of large hydrocarbons, up to virgin naphthas and heavy gasoils. The same approach, applied to oxidation and combustion process, is based on the generation of a detailed primary mechanism and then on the optimization of the kinetic parameters within a highly generic mechanism (similar, for example, to the Shell model (see Chap. 2)).

- *EXGAS* (Côme et al. 1997). The main specificity of this publicly distributed system is the use of the most comprehensive reaction class database and the large choice given to the user for mechanism tailoring. EXGAS has already been used to generate detailed mechanisms for alkanes (from C_4 to C_{16}), alkenes, cycloalkanes, ethers (Glaude et al. 2000), alcohols (Moss et al. 2008) and methylesters up to C_{19} (e.g. Hakka et al. 2010; Herbinet et al. 2011) which have been validated under a wide range of conditions. The mechanisms are composed of a comprehensive detailed primary mechanism, a lumped secondary mechanism and a C_0–C_2 (which can be supplemented with C_3–C_8 reactions) base mechanism.

- *RMG* (Green et al. 2001; Van Geem et al. 2006). This system distinguishes itself with a unique "generate and test" algorithm which generates a fundamental mechanistic step, estimates the rate constants and then, using a set of predefined physical conditions and cutoff criteria, determines "on-the-fly" whether the reaction should be included in the final mechanism (Susnow et al. 1997). The RMG system is also the only publicly distributed automatic generator of pressure-dependent reaction networks (Matheu et al. 2001; Allen et al. 2012).

- *REACTION* (Blurock 1995; Moreac et al. 2006). The main specificity of this last system is to use the concept of pathways instead of an exhaustive application of reaction classes. The result is the creation of mechanisms similar to those generated by hand. This system was also the first to represent the fundamental chemical information needed for generation solely in external databases independent of the generalized generation engine, so that the chemical information can be updated without modifying or recompiling the software.

3.2.1 *MAMOX++*

In the MAMOX++ and MAMA systems, detailed primary mechanisms are generated and then the parameters within a highly generic mechanism (similar, for example, to the Shell model) are optimized. For the further oxidation of the generated lumped products a hierarchical library of validated lumped mechanisms
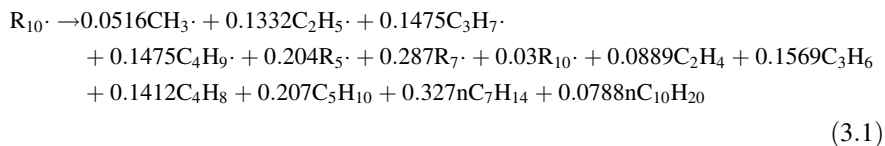
is searched and the corresponding lumped reactions used (Ranzi et al. 2005). The MAMA system in the SPYRO system (Dente et al. 1979) was the first to automatically generate mechanisms for the pyrolysis of hydrocarbons.

3.2.1.1 Lumping Procedure

The philosophy of the species lumping procedure is to produce a highly lumped primary mechanism derived from a comprehensive detailed primary mechanism. The generated lumped mechanism looks very similar to the Shell model of Halstead et al. (1975) and later derivations from Cox and Cole (1985), Hu and Keck (1987) and Cowart et al. (1991) in terms of the autoignition behavior of the fuel. They do however differ fundamentally in how the parameters are derived. Cowart et al. (1991) optimized the model for a particular fuel by adjusting the isomerization reaction of the alkyl peroxy radicals on the basis of engine data. The approach of MAMOX is to base the reduction on a detailed reaction mechanism. From the detailed mechanism, a set of (highly) lumped species and corresponding reactions are defined. The rate parameters are then optimized based on the generated detailed mechanism. The lumping rules are based on the primary product distributions predicted by the detailed model (Ranzi et al. 1995a, 2005).

3.2.1.2 Reduced Pyrolysis Mechanisms Using Steady State Approximation

The basic hypothesis for reduction under pyrolysis conditions is that the radicals larger than C_4 (μ-radicals) can only isomerize or decompose without significant interactions with the process mixture. In contrast, when generalizing this to oxidation reactions, it is necessary to take into account the fact that large intermediate radicals can interact with O_2. The alkyl radicals obtained from the hydrocarbon fuel after a single hydrogen atom abstraction are lumped into one lumped species, $R_n\cdot$, where n is the number of carbon atoms. For example, in the n-decane mechanism the 5 n-decyl radicals of the detailed mechanism would be lumped into a single $R_{10}\cdot$ species in the reduced mechanism. The sub-mechanism of all the isomerization and β-decomposition reactions is written explicitly in the detailed mechanism. Using this detailed mechanism, the (linear) system of continuity equations is solved under the steady state approximation at a given temperature to reduce the overall submechanism to one global reaction. For example, in the n-decane mechanism, the solution is (Ranzi et al. 2005):

$$R_{10}\cdot \rightarrow 0.0516CH_3\cdot + 0.1332C_2H_5\cdot + 0.1475C_3H_7\cdot$$
$$+ 0.1475C_4H_9\cdot + 0.204R_5\cdot + 0.287R_7\cdot + 0.03R_{10}\cdot + 0.0889C_2H_4 + 0.1569C_3H_6$$
$$+ 0.1412C_4H_8 + 0.207C_5H_{10} + 0.327nC_7H_{14} + 0.0788nC_{10}H_{20}$$

$$(3.1)$$

There are three key assumptions in this approach:

1. Radicals larger than C_4 can only isomerize or decompose without significant interactions with the process mixture, but do react with oxygen to form peroxy radicals.
2. There is only marginal importance of large radical recombinations and hydrogen atom abstractions. These are bimolecular reactions competing with fast unimolecular reactions.
3. The overall decomposition and radical distribution is temperature independent. Default temperature is assumed to be 1000 K.

The lack of interaction means that this sub-mechanism can be treated independently of other sub-mechanisms within the total pyrolysis mechanism. Not including recombination reactions affects the solution in two ways. First, isomerization and β-decomposition in the forward direction are unimolecular reactions and thus their solution under the steady state approximation is linear. However, recombination reactions are bimolecular. This would bring in non-linear terms. In addition, not including recombination reactions further "decouples" the distribution dependence with smaller radicals. This means the solution for larger radicals does not involve the distribution of the smaller radicals. Using the steady state approximation is temperature dependent. However, empirical evidence states that the radicals mainly decompose at temperature close to 1000 K and derived distributions are relatively temperature independent.

The purpose of the steady state approximation (see Chap. 18 for more details) is to reduce the complexity of the differential equations that must be numerically solved to determine the behavior of the fuel in a combustion process. In the simplest case, if one is interested in the time-dependent solution of the homogeneous adiabatic oxidation of a fuel, the system of reactions is translated into a system of differential equations. For example, the single β-decomposition reaction:

$$1\text{-}C_{10}H_{21}\cdot \rightleftarrows C_2H_4 + 1\text{-}C_8H_{17}\cdot \tag{3.2}$$

contributes to the following to the production and depletion of $1\text{-}C_{10}H_{21}\cdot$ radical:

$$\frac{d[1\text{-}C_{10}H_{21}\cdot]}{dt} = -k_f^D[1\text{-}C_{10}H_{21}] + k_r^D[C_2H_4][1\text{-}C_8H_{17}\cdot] \tag{3.3}$$

Note that in the forward direction there is a linear dependence on concentration. There is only one reactant, $1\text{-}C_{10}H_{21}\cdot$ radical. However, in the reverse direction, the radical addition, there is a non-linear dependence on concentration. There are two reactants, C_2H_4 and $1\text{-}C_8H_{17}\cdot$ radical. If the radical addition were considered insignificant relative to the β-decomposition, the second term could be neglected (equal to zero) and the dependence would be linear. In general, all β-decompositions (forward direction) contribute a linear term and all recombinations and radical addition contribute a non-linear term. If all recombination reactions were neglected (assumption number 2 above), then the system of equations describing all β-decompositions would be linear. Isomerizations are also represented by linear

equations. A single species isomerises to another and both the forward and reverse contributions to the differential equation are linear.

In general, the sum of all contributions from β-decomposition reaction D_j^i (the jth decomposition of species i), with rate constant $k_{i,j}^D$ and contribution from isomerisation reaction I_j^i (isomerization of species i to j), with rate constant $k_{i,j}^I$:

$$\frac{dX_i}{dt} = \sum_{D_j^i} k_{i,j}^D X_j + \sum_{I_j^i} k_{i,j}^I X_j \tag{3.4}$$

If a species i is in steady state, the concentration of that species does not change over time, i.e. $d[X_i]/dt = 0$. This transforms the set of linear differential equations to a set of linear algebraic equations. For a mechanism for a C_n alkane fuel, there are m alkyl radicals with n carbon atoms where $m = n/2$ (where n is even) or $m = n/2 + 1$ (where n is odd). This means there are m equations of the form of Eq. (3.4). These m algebraic equations can be used to represent the steady state concentrations of the radicals containing C_n carbon atoms. In the mechanism, this is represented as $R_n\cdot$. Successively solving the entire system allows the writing of the total β-decomposition of the $R_n\cdot$ radical in a form such as that shown in Eq. (3.1).

For the n-decane system, for example, there are five algebraic equations of the form of Eq. (3.4), one for each of the 5 C_{10} alkyl radicals. These can be used to derive their distribution in terms of the remaining alkyl radicals (those having less than 10 carbon atoms) and alkenes. The result is the distribution of the 5 alkyl radicals making up the lumped species, $R_{10}\cdot$. Table 3.1 empirically shows the relative temperature independence of the distribution for these 5 alkyl radicals. The distribution of species derived using the steady state procedure is actually temperature dependent.

One of the key assumptions MAMOX is that the temperature dependence of the reaction classes involved in lumping is minimal. The distribution derived at 1000 K is valid over a wide range of temperatures for the oxidation of hydrocarbons. The justification (Ranzi et al. 2005) is given in examining, for example, the product distribution of isomerization and β-decomposition of n-decyl radicals derived from hydrogen atom abstraction from n-decane (Table 3.1). The largest deviations stem from the primary 1-$C_{10}H_{21}\cdot$ radical and its subsequent decomposition to ethylene and a radical. However, Ranzi et al. (2005) make the further justification that above 1200 K the life-times of these radicals are short (10^{-8} s).

3.2.1.3 Reduced Oxidation Mechanisms

Reduction of oxidation mechanisms, for example for the low-temperature oxidation chemistry of hydrocarbons, does not satisfy the conditions under which pyrolysis mechanisms were reduced. Under oxidative conditions the sub-mechanisms of the reactions with oxygen are a significant "external interaction" and

Table 3.1 This table illustrates the temperature effect on the *n*-decyl radicals and the product distributions

	800 K	Rate constant	1200 K	Ave	σ	σ/Ave (%)
Relative concentrations (%)						
1-$C_{10}H_{21}\cdot$	4.75	8.50	13.30	8.9	4.29	48
2-$C_{10}H_{21}\cdot$	24.17	24.81	26.09	25.0	0.98	4
3-$C_{10}H_{21}\cdot$	23.62	22.33	21.14	22.4	1.24	6
4-$C_{10}H_{21}\cdot$	23.41	18.55	18.55	21.0	2.43	12
5-$C_{10}H_{21}\cdot$	24.05	23.32	20.92	22.8	1.64	7
Product distribution						
$C_9H_{18} + CH_3\cdot$	0.042	0.052	0.059	0.05	0.008	16
$C_8H_{16} + C_2H_5\cdot$	0.148	0.133	0.119	0.13	0.014	11
$C_7H_{14} + C_3H_7\cdot$	0.152	0.148	0.135	0.14	0.009	6
$C_6H_{12} + C_4H_9\cdot$	0.152	0.148	0.135	0.14	0.009	6
$C_5H_{10} + C_5H_{11}\cdot$	0.148	0.133	0.119	0.13	0.014	11
$C_4H_8 + C_6H_{13}\cdot$	0.149	0.141	0.136	0.14	0.007	5
$C_3H_6 + C_7H_{15}\cdot$	0.153	0.157	0.168	0.16	0.008	5
$C_2H_4 + C_8H_{17}\cdot$	0.056	0.089	0.130	0.09	0.037	40

The average (Ave) and standard deviation (σ) is over the three temperatures

cannot be neglected. In this case, another approach must be taken if the same degree of lumping is to be achieved. Instead of solving a set of linear equations as in the pyrolysis case, an optimization is used.

Anologous to the SHELL model, the major oxidation components are lumped (Ranzi et al. 1995a, b). In addition to $R_n\cdot$, where n is the number of carbon atoms in the alkyl species, the oxygenated species of the low-temperature chemistry are lumped: alkylperoxy radicals ($R_nOO\cdot$), hydroperoxyalkyl radicals ($\cdot Q_nOOH$), hydroperoxyalkylperoxy radicals ($\cdot OOQ_nOOH$), cyclic ethers ($ETER_n$) and keto-hydroperoxides (OQ_nOOH). The branching agents and products of the side chains of the low-temperature chemistry are alkenes including n or less carbon atoms, as well as aldehydes, ketones and aldehyde radicals with 3 or less carbon atoms. The lumped primary mechanism for the oxidation of *n*-decane is shown in Table 3.2.

Once again, the parameters of this low-temperature primary mechanism are optimized relative to the generated full detailed mechanism. The key to the optimization is to analyze the initial "cumulative selectivity" of the lumped species using the detailed mechanism (see Fig. 3.3). The kinetic parameters of the lumped primary mechanism are optimized using non-linear regression analysis. The difference between the cumulative selectivity predictions at each temperature of the lumped mechanism and the detailed mechanism are analyzed. The parameters of the lumped species are optimized until the squares of the differences are minimized.

Table 3.2 Primary lumped mechanism for the oxidation of n-decane

Reactions	A	Ea/R
$n\text{-}C_{10}H_{22} \rightarrow 0.09CH_3\cdot +0.26C_2H_5\cdot + 0.26C_3H_7\cdot + 0.26C_4H_9\cdot + 0.39\ R5\cdot + 0.5933\ R7\cdot + 0.1466\ R_{10}\cdot$	0.35×10^{18}	40750
$R\cdot + n\text{-}C_{10}H_{22} \rightarrow RH + R_{10}\cdot$	6 primary H-atoms and 16 secondary H atoms	
$O_2 \rightarrow HO_2\cdot$	0.23×10^{12}	22500
$OH\cdot \rightarrow H_2O$	0.53×10^{11}	850
$H\cdot \rightarrow H_2$	0.29×10^{12}	4000
… (see Ranzi et al. 2005)		
$R_{10}\cdot + O_2 \rightarrow n\text{-}C_{10}H_{20} + HO_2\cdot$	0.50×10^{9}	1750
$R_{10}\cdot \rightarrow 0.0516CH_3\cdot + 0.1332C_2H_5\cdot + 0.1475C_3H_7\cdot + 0.1475C_4H_9\cdot + 0.204R_5\cdot + 0.287R_7\cdot + 0.03R_{10}\cdot + 0.0889C_2H_4 + 0.1569C_3H_6 + 0.1412C_4H_8 + 0.207C_5H_{10} + 0.327nC_7H_{14} + 0.0788nC_{10}H_{20}$	0.35×10^{14}	15100
$R_{10}\cdot + O_2 \rightarrow R_{10}OO\cdot$	0.20×10^{10}	0
$R_{10}OO\cdot \rightarrow O_2 + R_{10}\cdot$	0.50×10^{14}	15600
$R_{10}OO\cdot \rightarrow \cdot Q_{10}OOH$	0.30×10^{13}	12100
$\cdot Q_{10}OOH \rightarrow R_{10}OO\cdot$	0.20×10^{11}	8100
$\cdot Q_{10}OOH \rightarrow ETHER_{10} + OH\cdot$	0.15×10^{11}	7100
$\cdot Q_{10}OOH \rightarrow n\text{-}C_{10}H_{20} + HO_2\cdot$	0.45×10^{13}	12100
$\cdot Q_{10}OOH \rightarrow 0.333CH_2O + 0.333CH_3CHO + 0.333C_2H_5CHO + 0.333n\text{-}C_7H_{14} + 0.333n\text{-}C_{10}H_{20} + OH\cdot$	0.20×10^{13}	11300
$O_2 + \cdot Q_{10}OOH \rightarrow \cdot OOQ_{10}OOH$	0.20×10^{10}	0
$\cdot OOQ_{10}OOH \rightarrow O_2 + \cdot Q_{10}OOH$	0.30×10^{15}	14600
$\cdot OOQ_{10}OOH \rightarrow OQ_{10}OOH + OH\cdot$	0.10×10^{13}	11600

Units are kmol, m, s, K. Lumped reactions for $n\text{-}C_{10}H_{20}$, $ETHER_{10}$ and $OQ_{10}OOH$ are part of the secondary mechanism (see Ranzi et al. 2005)
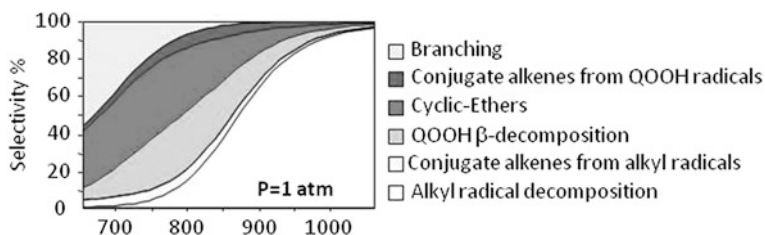
Fig. 3.3 The initial cumulative selectivities of the primary lumped species of *n*-decane oxidation at one atmosphere

3.2.2 EXGAS

Since the first attempts to model the oxidation of C_4–C_8 alkanes (Côme et al. 1997; Warth et al. 1998), the EXGAS program, which is written in Pascal, has been extensively used to produce detailed kinetic models for the oxidation of a wide range of hydrocarbons:

- linear and branched alkanes up to C_{16} (Buda et al. 2005; Biet et al. 2008; Herbinet et al. 2012),
- linear alkenes from C_3 to C_7 (Heyberger et al. 2001; Touchard et al. 2005; Bounaceur et al. 2009),
- cycloalkanes (Buda et al. 2006; Sirjean et al. 2007; Pousse et al. 2010).

Recent developments on oxygenated reactants are presented in Chap. 4. Software EXGAS-ALKANES-ESTERS automatically generates detailed kinetic mechanisms for the oxidation of linear and branched alkanes, and linear methyl esters and is freely available for academic researchers (valérie.warth@ensic.inpl-nancy.fr). Note also that the development of an EXGAS version dedicated to alkylbenzenes is in progress.

3.2.2.1 Notation

The external notation used to transfer the chemical formulae of the reactants and of the primary species between the user and the computer is a one-dimensional notation (Côme et al. 1984). This "linear notation" is very close to the semi-developed notation in the case of non-cyclic compounds. For the ease of use by modelers, this notation is non-ambiguous, but also non-canonical as shown in Fig. 3.4 for *n*-dodecane. That means that a same species can be represented by different notations. In the internal notation, the two types of chemical acyclic species (molecules and free radicals), which are involved in this program, are represented by a tree-like structure (see example in Fig. 3.4) on which is applied an algorithm of canonicity. More details about these internal and external
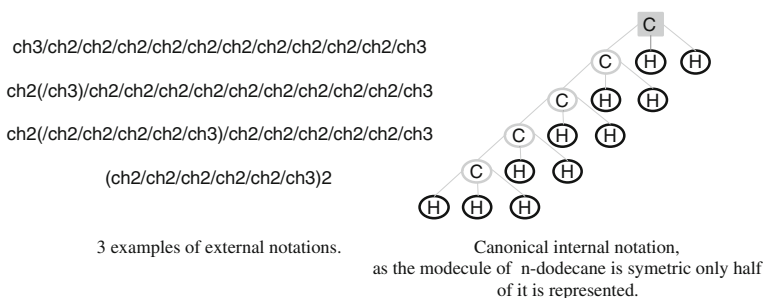
ch3/ch2/ch2/ch2/ch2/ch2/ch2/ch2/ch2/ch2/ch2/ch3

ch2(/ch3)/ch2/ch2/ch2/ch2/ch2/ch2/ch2/ch2/ch3

ch2(/ch2/ch2/ch2/ch2/ch3)/ch2/ch2/ch2/ch2/ch3

(ch2/ch2/ch2/ch2/ch2/ch3)2

3 examples of external notations.

Canonical internal notation,
as the modecule of n-dodecane is symetric only half
of it is represented.

Fig. 3.4 External and internal notation in the case of *n*-dodecane

notations, and the notations used for cyclic molecules can be found in (Warth et al. 2000).

3.2.2.2 General Structure of an EXGAS Model

As presented in Fig. 3.5, a model generated by EXGAS is composed of three parts: a C_0–C_2 base mechanism including all the reactions involving radicals or molecules containing less than three carbon atoms, a comprehensive primary mechanism, and a lumped secondary mechanism, containing reactions consuming the molecular products of the primary mechanism, which do not react in the reaction bases.

Thermochemical data for molecules or radicals are automatically calculated and stored as 14 polynomial coefficients, according to the CHEMKIN formalism (Kee
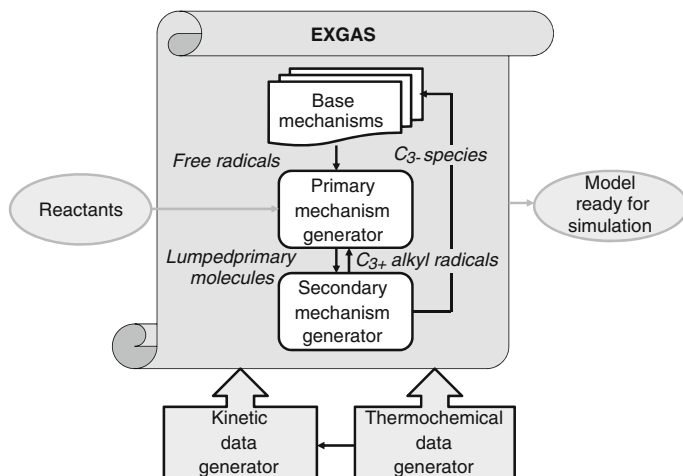

Fig. 3.5 General structure of the EXGAS system (adapted from Warth et al. 1998)

et al. 1993). These data are calculated using the software THERGAS (Muller et al. 1995) based on the group and bond additivity methods proposed by Benson (1976) (see Chap. 20).

The kinetic data of the reactions included in the primary or secondary mechanisms are also automatically provided: they are either calculated using thermochemical kinetic methods (Warth et al. 1998) or estimated using a wide range of correlations (Warth et al. 1998; Heyberger et al. 2001; Buda et al. 2005; Touchard et al. 2005; Biet et al. 2008; Glaude et al. 2010).

3.2.2.3 The Base Mechanisms

The C_0–C_2 base mechanism used by EXGAS was initially written by Barbé et al. (1995). Since then it has been continuously up-dated (see latest revision by Cord et al. (2012)). The pressure dependent rate constants follow the formalism proposed by Troe (1974) and efficiency coefficients have been included. This base mechanism can easily be completed to also consider the reactions of C_3–C_5 polyunsaturated hydrocarbons (Gueniche et al. 2009) and those of small aromatic compounds such as benzene, toluene or ethylbenzene (Husson et al. 2013).

3.2.2.4 The Primary Mechanism Generation

Figure 3.6 presents the structure of the algorithm (or generator engine) used for the generation of primary mechanism for alkanes and alkenes as well as the involved reaction classes. At the beginning, the species pools only contains the initial reacting hydrocarbons, which can be a single molecule or a mixtures, oxygen and the radicals present in the C_0–C_2 base mechanism. Then the first C_{2+} radicals are created by all the possible initiations from the molecular reactants and enter the species pool. In a second step, all radicals present in the species pool are submitted to the reactions present in the propagation loop; the radicals of the C_0–C_2 base mechanism react only by addition to the double bond or hydrogen atom abstraction with the reacting hydrocarbon. Each new created radical is to its turn included to the species pool and the algorithm is terminated when no new radical is created in the propagation loop. As the additions of carbon atom containing radicals is considered as a reaction class in the case of alkenes, the user needs then to specify a maximum species size to avoid the creation of an infinite propagation loop. In a final step, all the radicals created by initiations and propagations should be submitted to termination steps.

To avoid unnecessarily long reaction mechanisms, the following simplifying rules are generally used. Three different classes of radicals have been identified by the $\beta\mu$ rules of Goldfinger-Letort-Niclause (Warth et al. 1998): β radicals which cannot decompose by unimolecular process (typical β free radicals are ·H, ·OH, or ·CH$_3$), μ radicals which can easily decompose by a unimolecular process involving the scission of a (C–C) or a (C–O) bond (typical μ free radicals are
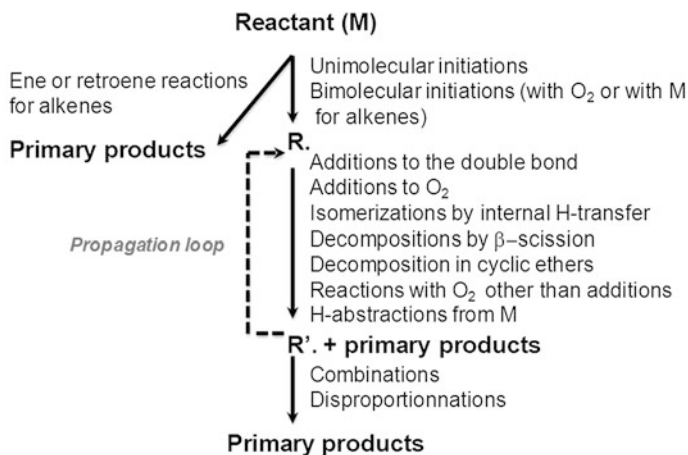
Reactant (M)

Ene or retroene reactions
for alkenes

Unimolecular initiations
Bimolecular initiations (with O_2 or with M
  for alkenes)

Primary products

R.

Additions to the double bond
Additions to O_2
Isomerizations by internal H-transfer
Decompositions by β–scission
Decomposition in cyclic ethers
Reactions with O_2 other than additions
H-abstractions from M

Propagation loop

R'. + primary products

Combinations
Disproportionnations

Primary products

Fig. 3.6 Algorithm of generation used to generate the primary mechanisms of alkanes and alkenes in EXGAS

n-C_3H_7· and s-C4H9·), $\beta\mu$ *radicals* which have a β behaviour at low temperatures and a μ behaviour at high temperature (typical $\beta\mu$ radicals are ·OOH, ·CHO, ·OCH$_3$, or ·OOCH$_3$). Therefore according to these classes, the radicals involved in bimolecular reactions (e.g. H-abstractions or termination steps) are mostly of β and $\beta\mu$ types. More details can be found in Warth et al. (1998).

It is made possible for hydroperoxyalkyl radicals (·QOOH) to undergo a second addition to oxygen yielding ·OOQOOH radicals. Note that the possibility of a third addition to oxygen is not considered. The detailed isomerizations and subsequent decompositions of ·OOQOOH radicals can be considered, but in most cases, only the direct formation of ·OH radicals and a globalized ketohydroperoxide is written. The rate of this global step for a given ·OOQOOH radicals is the sum of the rates of all the possible isomerizations of this radical (Glaude et al. 2000).

Note also that the considered reaction classes and the level of simplification (e.g. types of radicals to be considered) can be chosen by the kineticist prior to the generation. For instance, for modelling a system above 1000 K, the generation of the addition to oxygen and the subsequent reactions could be discarded. Also breaking C–H bonds during β-scission decompositions can be omitted if these reactions can be neglected under the studied conditions (Glaude et al. 2000).

3.2.2.5 The Secondary Mechanism Generation

In most cases, the molecular products formed in the primary mechanism are lumped in order to reduce the size of the model. Lumping consists here in gathering the molecular products having the same global formulae and the same functional groups into one generic species: as an example, all isomers of linear

dodecene, which are primary products obtained during the oxidation of *n*-dodecane, are lumped as a single species: $C_{12}H_{24}$. This is very similar to what is used in MAMOX++. However when needed some important primary products (e.g. cyclic ethers) can be considered individually (Herbinet et al. 2012).

Secondary reactions are written for every type of molecular products: hydroperoxides, alkenes, cyclic ethers, aldehydes, alkanes, ketones, alcohols. However to avoid an explosion of the size of the mechanisms, these reactions are written in a global way in order to promote the formation of species which are already included in the primary mechanism and in the C_0–C_2 reaction base (Biet et al. 2008; Glaude et al. 2010). The kinetic data of the reactions of primary products generated by EXGAS are those of the first involved reaction: O–OH bond breaking for hydroperoxides, hydrogen atom abstractions for cyclic ethers, aldehydes, alkanes, ketones, and alcohols, and finally radical additions and retro-ene decompositions for alkenes.

Table 3.3 presents the lumped reactions automatically written by EXGAS for the consumption of lumped dodecene. Considering the formation of alkyl radicals, such as $\cdot C_{12}H_{25}$ or $\cdot C_{10}H_{21}$, the reactions of which are taken into account in details in the primary mechanism, allows a more accurate representation of the involved decomposition channels. Secondary molecules such as $C_{10}H_{21}CHO$ or $C_{12}H_{24}O$-oxirane react again in the secondary mechanism, but with reactions of aldehydes and cyclic ethers, respectively (Biet et al. 2008).

Table 3.3 Lumped reactions written by EXGAS for the oxidation of dodecene

Other species involved	Products	Rate constant
	Additions	
H-atoms	$\cdot C_{12}H_{25}$	$1.32.10^{13}$ exp($-785/T$) $+ 1.32.10^{13}$ exp($-1640/T$)
$\cdot OH$	$HCHO + \cdot C_{11}H_{23}$	1.4×10^{12} exp($520/T$)
	$CH_3 + C_{10}H_{21}CHO$	1.4×10^{12} exp($520/T$)
O-atoms	$CH_2CO + H$ $+ \cdot C_{10}H_{21}$	6.0×10^4 $T^{2.56}$ exp($770/T$)
$\cdot CH_3$	$C_4H_8 + \cdot C_9H_{19}$	1.7×10^{11} exp($3720/T$)
	$C_3H_6 + \cdot C_{10}H_{21}$	9.6×10^{10} exp($4030/T$)
$\cdot OOH$	$\cdot OH + C_{12}H_{24}O$-oxirane	1.0×10^{12} exp($7150/T$)
	H-abstractions	
Small radicals ($\cdot X$), such β or $\beta\mu$ radicals	$XH + C_4H_6$ $+ \cdot C_8H_{17}$	As for the abstraction of 2 secondary allylic, 3 primary alkylic and 16 secondary alkylic hydrogen atoms from initial alkene as proposed by Buda et al. (2005), Heyberger et al. (2001))
	Retro-ene decompositions	
–	$C_3H_6 + C_9H_{18}$	8.0×10^{12} exp($-28400/T$)

Rate constants are given in cm^3, s^{-1}, mol units

3.2.3 RMG

The RMG open-source software package has its conceptual origins in the NetGen software developed by Broadbelt et al. (1994), which uses an exhaustive generation technique (see Sect. 3.1.2.3) of fundamental reaction steps (Green 2007). RMG is unusual in that it tests each generated species using a rate-based criterion "on-the-fly" (meaning during the generation process) to see whether it should be included in the final mechanism (Susnow et al. 1997). So while all reactions are considered, with the numerical rate-based tests only significant reactions filter through. This is a classic "generate and test" algorithm from classical artificial intelligence (Nilsson 1982). The success of this test is highly dependent on the accuracy of the derived thermodynamics and rate constants.

Another unusual feature of RMG is that it automatically identifies chemically-activated reaction sub-networks, using a similar strategy to that used for ordinary thermal reaction networks (Allen et al. 2012; Matheu et al. 2003). The pressure-dependent rate coefficients for all the reaction pathways inside each sub-network are computed using approximations to the master equation (master equation is described in Chap. 21). This special capability of RMG is particularly helpful for high-temperature low-pressure systems, where more than half of the important reactions may have significantly pressure dependent rate-coefficients. Most mechanisms generated in other ways omit many of these chemically-activated reaction pathways.

The reaction and species database and the methods used in the generation process also set this system apart from the others. The rate estimation rules are stored in an editable database, rather than being hard-coded into the software, conceptually similar to the approach of Blurock (1995) (see Sect. 3.2.4). This makes the software more maintainable, and it is easier to update rate estimates, or to add new ones. Similar to NetGen and related XMG software (Grenda et al. 2003; Matheu et al. 2003), some of the thermochemistry is computed on the fly during model generation by spawning quantum chemistry jobs (Magoon and Green 2013).

The RMG software was originally written by Song et al. (2004). The first published demonstration of its capabilities was (Van Geem et al. 2006). RMG currently includes free radical and concerted-reaction chemistry for molecules containing C, H, O, and S atoms. The current version does not consider ions or photochemistry, but optionally includes solvent effects on thermochemistry and on some rate coefficients (Jalan et al. 2013), so it can predict reactions in liquid phase. For a recent demonstration of its capability to predict complicated chemistry see (Harper et al. 2011). The latest version is available at http://rmg.sourceforge.net.

3.2.3.1 Rate Based Generation Algorithm: Network Expansion and Termination Criteria

At any given point in the algorithm, the species pool is divided into two sets, "*reacted*" and "*unreacted*" species (see Fig. 3.7). The species in the reacted set
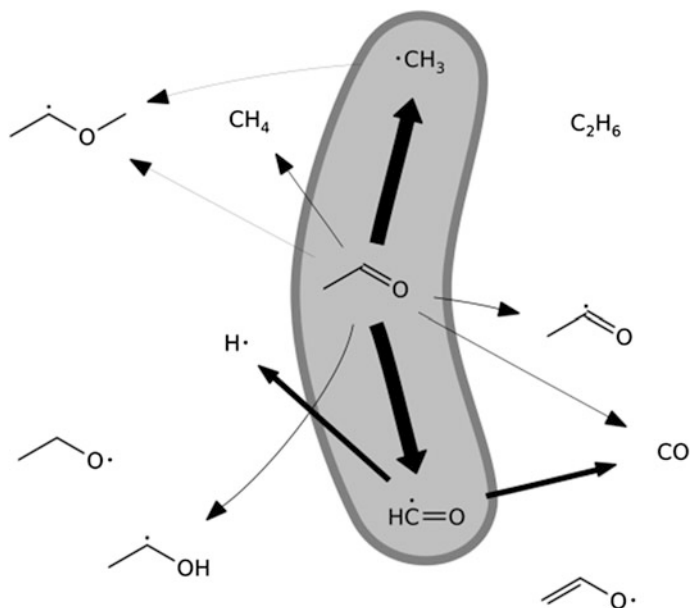
Fig. 3.7 Species pool connected by a reaction network in the case of the pyrolysis of acetaldehyde. The species pool is divided into two regions, reacted and unreacted

have been reacted with all possible allowed reaction classes. The unreacted species are products of the reactions of the reacted species. But the reaction classes have not yet been applied to the unreacted species.

The basis of the algorithm is to expand a network of reacted species one species at a time. The species chosen to be expanded is the unreacted species having the highest rate of formation. This is the species that is drawing out the most flux from the reacted network. The reaction network is deemed complete when this rate of formation is less than a given criteria. The reaction network is self-consistent and the perturbation of any additional reactions to outside this network is small, i.e. a minimum of flux leaves the network. In this philosophy of the generation the flow from a reacted species inside the network to a unreacted species outside the network is analyzed (Susnow et al. 1997). The unreacted species lie on the "edge" of the reaction network. The further reactions of the unreacted species could produce an expanded network involving "unknown" species. The network is only generated when the flux to the unreacted species at the edge of the reacted network is large enough.

During the analysis, the network is converted to a reactive system where each of the reactions is converted to a differential equation with respect to time as shown in Chap. 2. The system of equations is solved at regular time intervals until a particular reactant conversion, X_A, is achieved at time τ. At each time step the total rate of production (see Eq. (2.3) of Chap. 2) of each species, R_i, is determined. For

the reacted species, this is the sum of the rates of formation minus the rates of consumption. Since their consumption reactions have not been formed yet, only the rate of formation is computed for the unreacted species in the network. The maximum rate of formation for each species, $R_{i,\max}$, is determined for all time, t, intervals ($0 < t < \tau$). Note that R_i for all unreacted species is positive. The largest $R_{i,\max}$ of the unreacted species is the next species to be added to the reacted species network. The entire reaction network consisting of reacted and unreacted species is expanded with the given reaction classes applied to this new reacted species.

The expansion of the network is terminated when all the maximum rates of formation, $R_{i,\max}$, of the unreacted species is under a threshold, $R_{\min}$ (Susnow et al. 1997). $R_{\min}$ represents a combination of a characteristic rate (e.g. the disappearance rate of the reactant) R_{char}, indicating how fast the system is evolving, and the desired precision level, $f_{\min}$:

$$R_{\min} = f_{\min} R_{\mathrm{char}} \tag{3.5}$$

One possible choice of R_{char} is the average conversion rate over time τ defined as:

$$R_{\mathrm{char}} = [C_{A0} - C_A(\tau)]/\tau = X_A C_{A0}/\tau \tag{3.6}$$

where C_{A0} is the initial concentration of the reactant and C_A is the concentration of the reactant at the particular conversion X_A. If the formation of all unreacted species on the edge of the reacted network of species is less than the tolerance $R_{\min}$, then it is deemed that there is no significant flux outside the reacted network. Thus the network is deemed complete and the generation algorithm terminates. Upon termination, the sum total of the "leaks" to the unreacted species gives a measure of the overall error due to incompleteness of the reaction network.

Dividing up the species into reacted and unreacted species simplifies the solution to the differential equations resulting from the reaction network: the reverse reaction to that producing a unreacted species is ignored. This decouples the computation of the differential equations:

$$\frac{\mathrm{d}C_j}{\mathrm{d}t} = f(C_j) \tag{3.7}$$

$$R_i = \frac{\mathrm{d}C_i}{\mathrm{d}t} = g(C_j) \tag{3.8}$$

Where j is over all "reacted" species and i is over all unreacted species. Both functions f Eq. (3.7) and g Eq. 3.8) are (algebraic) polynomial functions in only reacted species. This means that the n differential equations Eq. (3.7), where n is the number of reacted species, is solved first for the set of reacted concentrations, C_j. This solution is then substituted into the algebraic expressions of the unreacted species Eq. (3.8) to determine each R_i.

The reacted network is solved independently of the unreacted network. This means the reacted network is considered a closed system and ignores the leaks into

the unreacted system. As the system nears completion, i.e. the leaks get smaller and smaller, the errors from this approximation become less significant. The independent calculation of each R_i is sufficient for the decision making process, i.e. to choose which species to expand next and whether the algorithm should terminate. The size of the set of reacted species is always considerably smaller that the size of the unreacted species set and hence represents a considerable computational savings.

As currently implemented in RMG this selection algorithm has been found to be very effective at automatically identifying the reaction pathways leading to major products and byproducts without including many unimportant species, but it sometimes omits low-flux sensitive reactions (e.g. degenerate branching steps important in ignition). Of course, the termination criteria of is not unique. The formulation in NetGen (Broadbelt et al. 1994), from which RMG is conceptually derived, of the termination criteria was based on carbon atom count (Broadbelt et al. 1994) or "rank" of the reaction (Bhore et al. 1990; Broadbelt et al. 1994).

3.2.3.2 Pressure Dependent Networks: Activated Species Algorithm

The "Activated Species Algorithm" (ASA) treats chemically or thermally activated species as though they were distinct species to produce the connectivity of the pressure dependent channel networks (Matheu et al. 2001; Matheu 2002; Allen et al. 2012). The general flow of the algorithm is the same as with rate-based generation algorithm described previously.

If a reaction produces a single product (see example in Fig. 3.8), then it is a candidate for the pressure dependent network. The product, C*, is added as an unreacted species. The reaction is added to the network with an initial rate constant corresponding to the high-pressure-limit, k^∞. Using k^∞ gives the maximum possible flux through the reaction channel to any possible species in the pressure-dependent network. If its flux of formation satisfied the screening criteria, an activated network containing the reactions of C^*, such as isomerizations, collisional quenching or the production of exit products is explored (Grenda et al. 2003).

For the reactions currently in the reaction list (the 3 reactions in the right side of Fig. 3.8), temperature and pressure dependent rate constants are calculated, k(T,P), using a modified version of CHEMDIS software (Bozzelli et al. 1997; Chang et al. 2000; Allen et al. 2012)). This code only gives an approximation (see discussion in Matheu et al. (2003) of the true pressure-dependent phenomenological rate coefficients (Miller and Klippenstein 2012) however CHEMDIS provides quick "on the fly" estimates of rate constants for complex, multiwell pressure dependent network using input quantities that are readily available in mechanism generation systems. In a post processing step, the computed rate coefficients, k(T,P), are fitted to Chebyshev, logP, or Arrhenius form (see more details in Chap. 19). The Arrhenius values obtained from this sort of fitting have no simple physical meaning, and change with pressure.
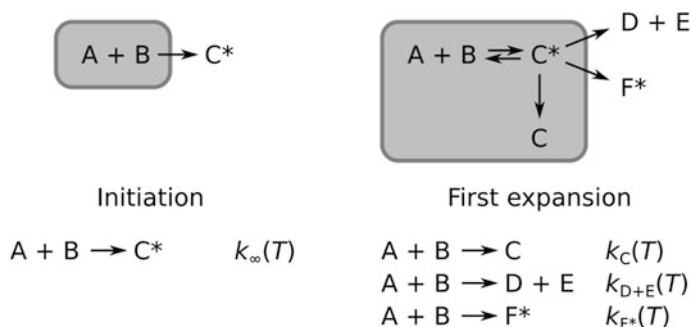
Fig. 3.8 Activated species algorithm: the reaction $A + B$ forms a single activated species, $C*$. The activated species is then expanded to additional activated species, D, E and $F*$. In addition the stabilized species, C, is added to the network

The reaction $A + B \rightarrow C*$ is then replaced by the reaction $A + B \rightarrow C$ with $k_C(T,P)$ calculated by CHEMDIS and D, E and F* are added as unreacted species. The criteria to determine whether the network is complete or not are similar to the general RMG criteria. It is dependent on how much flux "leaks" (R_{leak}), out of the network. In the case of pressure-dependent networks, this is the total flux to the activated species in the unreacted species list which is screened using temperature and pressure dependent rate constants. In the case of Fig. 3.8:

$$R_{leak} = (k_{D+E}(T) + k_{F*}(T))[A][B] \tag{3.9}$$

For the pressure-dependent networks the characteristic rates in Eq. (3.5) have another form (for comparison see Eq. (3.6)) describing the total "input rate" to the pressure dependent network. For chemical activation (as in Fig. 3.8), the characteristic rate takes the form:

$$R_{char} = k_{inp}^{\infty}[A][B] \tag{3.10}$$

For dissociation reactions, the characteristic rate is the fraction of the collisions between the reactant, C and the bath gas which produce $[C*(E)]$ at energies above the lowest barrier to dissociation (for a full explanation of terms see (Chang et al. 2000)):

$$R_{char} = \int_{E_0}^{E_{max}} k^s(T)[M] \frac{\rho_C(E)e^{-E/k_B T}}{Q_C(T)} dE \, [C] \tag{3.11}$$

Briefly, ρ_c is the density of states for C at energy E (calculated using thermodynamic estimates, see (Bozzelli et al. 1997), Q_C is the partition function for C).

If R_{leak} is above the threshold, the isomer corresponding to the largest element of the sum is selected as the next candidate for exploration. If it is $F*$, the reaction

$A + B \rightarrow F$ (F being the stabilized species corresponding to F^*) is added to the reaction list with using $k_F(T)$ calculated by CHEMDIS, while the formation of $D + E$ is not included in the model, i.e. $D + E$ remains on the edge.

3.2.4 Reaction

The REACTION system (Blurock 1995; Moreac et al. 2006) is the first "data-driven" automatic generation system used in combustion. The primary design strategy is to have the hard-coded generation "engine" as small and as general as possible and let the external database guide the generation procedure. The standard reaction classes, for example from Curran et al. (1998), are in the form of an external database in 2D-graphical representations of the reactive center and important surrounding functional groups (Blurock 2004a, b). In addition, the generation strategy, i.e. how the reaction classes are applied, is not fixed as in other systems. This is accomplished through the concept of reaction pathways, i.e. a sequence of reaction classes, to generate mechanisms. This is in line with a main goal of REACTION which is to mimic how the combustion modeler thinks and generates a mechanism. The use of pathways instead of recursive use of a pool of reaction classes is closer to how a modeler would produce a mechanism.

REACTION stems from earlier work on the RETROSYN computer program (Blurock, 1990) for computer aided organic synthesis (CAOS). Both CAOS and automatic generators use reaction classes. The main difference is that in CAOS the modeler starts with the molecule to be synthesized and uses the reaction classes retro-synthetically (Corey and Wipke 1969; Wipke and Howe 1977), i.e. in reverse, to derive the starting reactants. The goal of RETROSYN was not to use "programmed" reaction classes, but to derive the reaction classes from an electronic database of synthetic organic reactions (Blurock 1990). The key to the process was to determine the maximal common subgraph (a special case of graph isomorphism) to determine the reactive center between the reactants and products. This derived database in the form of 2D-graphical structures was used to perform retro-synthetic analysis.

3.2.4.1 Reaction Patterns

Though, there is no strict definition of reaction class, a single reaction class describes a single type of reactivity. But all reactivity within a reaction class does not need to have the same rate constant. The reactive environment beyond, for example, the reactive center, could influence the rates. For example, a common distinction within a single reaction class is the effect of primary, secondary and tertiary carbon atoms. If a reaction class is to be described with 2D-graphical structures each of these different environments need be included. In REACTION this is done with reaction patterns (Blurock 1995; Ratkiewicz and Truong 2006)

where each individual distinguishing chemical environment leading to a different rate constant is represented as a set of reactant and product 2D graphical substructures.

For example, the hydrogen atom abstraction reaction class encompasses not only a wide range of abstractors, but also a range of types of hydrogen atoms. In the description of the reaction class by Curran et al. (1998), 11 radical abstractors are applied to hydrogen atoms on primary, secondary and tertiary carbon atoms. In order to describe these reactions with graphical substructures 33 reaction patterns are needed corresponding to the 11 abstractors and 3 types of carbon atom. The primary, secondary and tertiary carbon atoms are represented as generic carbon graphs. Matching is made through graphical isomorphism and atom-to-atom correspondence as described in Sect. 3.1.3. To apply this reaction pattern, the carbon structure is matched within the carbon and the hydroxyl radical is matched to the reactants in the species pool.

3.2.4.2 Pathways

As outlined in Chap. 2, a detailed mechanism is not just a collection of reactions, but has a distinct structure. One such structure is the reactive pathway. This is a linear sequence of reactions from an initial reactant to products. When creating a reaction mechanism, the modeler often starts with an initial species, possibly the fuel, and applies an initial reaction class. The products are then examined and the next reaction class is applied. One classic pathway that is built up in this way in combustion is the low temperature pathway (see Sect. 2.2.3) from fuel through to branching agents. This is an efficient means of building up a (sub)mechanism ensuring that the products of all reactions are consumed by other reactions.

REACTION differs from other automatic generators in that instead of applying recursively a pool of reaction classes, the reaction classes are arranged in pathways, i.e. a linear sequence of sets of reaction classes. The primary reasons for generating in this way are:

- A more controlled form a generation which inhibits the combinatorial explosion of possible reactions.
- It mimics more closely the way a modeler thinks and builds a complex detailed mechanisms.
- It provides a means of introducing a generation strategy without hard-coding the strategy in the generation engine.

The main difference in the generation procedure is that instead of continually adding newly generated species in the species pool, the species pool is initialized after every step with only newly generated species. This means that the reaction classes of the current step are only applied to the products of the last step. In designing a pathway, those products not consumed by one of the steps should be consumed by another submechanism.

A systematic comparison of the comprehensive *n*-hexadecane mechanism as generated by REACTION and that produced by Westbrook et al. (2009) yielded very similar mechanisms. Both mechanisms included all *n*-alkane mechanisms up until hexadecane. For the species with eight or more carbon atoms the two mechanisms were exactly the same for almost all the classes. They differed only when hand-generated mechanism did extensive lumping and in a few cases where the hand-generated did not include all combinations of reaction class applications.

In REACTION the reaction path is applied to a seed molecule and produces a submechanism. For the *n*-decane mechanism (Moreac et al. 2006), the *n*-pentane to *n*-decane seeds molecules were applied to 8 pathways to produce 48 generated submechanisms. For the *n*-tetradecane mechanism, *n*-pentane to *n*-hexadecane seed molecules were applied to 10 reaction pathways to produce 160 submechanisms. These included high and low temperature chemistry.

The important consideration in the design of pathways is that the products which are produced in the last step are the initial molecules of another submechanism. In general the REACTION pathways have been designed so that the final products have only one functional group: a simple radical, aldehyde, ketone and simple alkene or alkyne. There are pathways which involve these species or have them as initial molecules.

3.2.4.3 Generated Submechanisms and Base Submechanism

A generated mechanism from REACTION consists of all the generated submechanisms combined with a literature base mechanism. The "communication" between the mechanisms occurs between the product molecules of one mechanism and reacting molecules in another.

For the generated mechanisms the 2D graphical structure is known. Thus combining these mechanisms to one mechanism is straightforward, one needs only use graph isomorphism to decide if the molecules are the same or not. Consequently, one can determine whether the reactions are repeated also.

As outlined in Chap. 2, the species in a literature mechanism for numerical calculations are just labels. The 2D-graphical structure is not needed. However, to combine a literature mechanism with the generated mechanism a correspondence table with the species name in the literature mechanism and the 2D-graphical structure is needed. REACTION has a database of standard molecules with corresponding 2D graphical information. The correspondence table needs only give the literature species name that is associated with the REACTION database species name. If a generated reaction and literature reaction are the same, the literature reaction is chosen for the combined mechanism.

Given this correspondence table and the literature mechanism (in CHEMKIN format), a final combined mechanism can be created. A tool is provided by the REACTION system to facilitate the creation of the correspondence table. It essentially detects which molecules are not consumed in a forward reaction by the generated mechanisms. For the generated *n*-decane mechanism (Moreac et al.

2006), a literature mechanism based on Hoyermann et al. (2004) was used. This mechanism included essentially C_2 compounds (species with up to two carbon atoms) and was supplemented with additional reactions from the literature (see (Moreac et al. 2006). The generated hexadecane mechanism was combined with the C_4 mechanism of Westbrook's *n*-hexadecane mechanism (Westbrook et al. 2009).

3.3 Concluding Summary

This chapter has drawn the main lines of automatic generation of combustion models, trying to enlighten some specificities of the four most advanced individual systems:

- MAMOX, with its particularly efficient extensive use of lumping,
- EXGAS, which is currently able to consider the widest range of reactive fuels and biofuels,
- RMG, with its unique rate-based "on-the-fly" screening of reaction and identification of chemically-activated reaction sub-networks,
- REACTION with its use of pathway.

References

Allen JW, Goldsmith CF, Green WH (2012) Automatic estimation of pressure-dependent rate coefficients. Phys Chem Chem Phys 14:1131–1155

Babai L, Luks EM (1983) Canonical labeling of graphs. In: Proceedings of the fifteenth annual ACM symposium on theory of computing. Presented at the STOC'83, Association of Computing Machines, New York, pp 171–183

Balaban AT (1985) Applications of graph theory in chemistry. J Chem Inf Comput Sci 25:334–343

Balaban AT (1995) Chemical graphs: looking back and glimpsing ahead. J Chem Inf Comput Sci 35:339–350

Barbé P, Battin-Leclerc F, Côme GM (1995) Experimental and modelling study of methane and ethane oxidation between 773 and 1573 K. J Chim Phys 92:1666–1692

Barnard JM (1993) Substructure searching methods: old and new. J Chem Inf Comput Sci 33:532–538

Benson SW (1976) Thermochemical kinetics, 2nd edn. Wiley, New York

Bhore N, Klein M, Bischoff K (1990) Species rank in reaction pathways—application of delplot analysis. Chem Eng Sci 45:2109–2116

Biet J, Hakka MH, Warth V et al (2008) Experimental and modeling study of the low-temperature oxidation of large alkanes. Energ Fuels 22:2258–2269

Blurock E (1990) Computer-aided synthesis design at Risc-Linz—automatic extraction and use of reaction classes. J Chem Inf Comput Sci 30:505–510

Blurock E (1995) Reaction—system for modeling chemical-reactions. J Chem Inf Comput Sci 35:607–616

Blurock ES (2004a) Detailed mechanism generation. 1. Generalized reactive properties as reaction class substructures. J Chem Inf Comput Sci 44:1336–1347

Blurock ES (2004b) Detailed mechanism generation. 2. Aldehydes, ketones, and olefins. J Chem Inf Comput Sci 44:1348–1357

Bounaceur R, Warth V, Sirjean B et al (2009) Influence of the position of the double bond on the autoignition of linear alkenes at low temperature. Proc Combust Inst 32:387–394

Bozzelli JW, Chang AY, Dean AM (1997) Molecular density of states from estimated vapor phase heat capacities. Int J Chem Kinet 29:161–170

Broadbelt LJ, Stark SM, Klein MT (1994) Computer generated pyrolysis modeling: on-the-fly generation of species, reactions, and rates. Ind Eng Chem Res 33:790–799

Buda F, Bounaceur R, Warth V et al (2005) Progress towards an unified detailed kinetic model for the autoignition of alkanes from C_4 to C_{10} between 600 and 1200 K. Combust Flame 142:170–186

Buda F, Heyberger B, Fournet R et al (2006) Modeling of the gas-phase oxidation of cyclohexane. Energ. Fuels 20:1450–1459

Chang AY, Bozzelli JW, Dean AM (2000) Kinetic analysis of complex chemical activation and unimolecular dissociation reactions using QRRK theory and the modified strong collision approximation. Phys Chem Chem Phys 214:1533–1568

Chen WL (2006) Chemoinformatics: past, present, and future. J Chem Inf Model 46:2230–2255

Côme GM, Muller C, Cunnin PY et al (1984) A linear chemical notation. Comput Chem 8:233–237

Côme GM, Warth V, Glaude PA et al (1997) Computer-aided design of gas-phase oxidation mechanisms—application to the modeling of n-heptane and iso-octane oxidation. Proc Combust Inst 26:755–762

Cord M, Sirjean B, Fournet R et al (2012) Improvement of the modeling of the low-temperature oxidation of *n*-butane: study of the primary reactions. J Phys Chem A 116:6142–6158

Corey EJ, Wipke WT (1969) Computer-assisted design of complex organic syntheses. Science 166:178–192

Cowart JS, Keck JC, Heywood JB et al (1991) Engine knock predictions using a fully-detailed and a reduced chemical kinetic mechanism. Proc Combust Inst 23:1055–1062

Cox R, Cole J (1985) Chemical aspects of the autoignition of hydrocarbon-air mixtures. Combust Flame 60:109–123

Curran HJ, Gaffuri P, Pitz WJ et al (1998) A comprehensive modeling study of n-heptane oxidation. Combust Flame 114:149–177

Dalby A, Nourse JG, Hounshell WD et al (1992) Description of several chemical structure file formats used by computer programs developed at molecular design limited. J Chem Inf Comput Sci 32:244–255

Dente M, Ranzi E, Goossens AG (1979) Detailed prediction of olefin yields from hydrocarbon pyrolysis through a fundamental simulation model (SPYRO). Comput Chem Eng 3:61–75

Dente M, Pierucci S, Ranzi E et al (1992) New improvements in modeling kinetic schemes for hydrocarbons pyrolysis reactors. Chem Eng Sci 47:2629–2634

Dugundji PDJ, Ugi PDI (1973) An algebraic model of constitutional chemistry as a basis for chemical computer programs. Computers in Chemistry, Fortschritte Der Chemischen Forschung. Springer. Berlin, pp 19–64

Engel T (2006) Basic overview of chemoinformatics. J Chem Inf Model 46:2267–2277

Engel T, Gasteiger J (2002) Chemical structure representation for information exchange. Online Inf Rev 26:139

Gillespie RJ, Robinson EA (2007) Gilbert N. Lewis and the chemical bond: the electron pair and the octet rule from 1916 to the present day. J Comput Chem 28:87–97

Glaude PA, Battin-Leclerc F, Fournet R et al (2000) Construction and simplification of a model of the oxidation of alkanes. Combust Flame 122:451–462

Glaude PA, Herbinet O, Bax S et al (2010) Modeling of the oxidation of methyl esters-validation for methyl hexanoate, methyl heptanoate, and methyl decanoate in a jet-stirred reactor. Combust Flame 157:2035–2050

Green WH (2007) Predictive kinetics: a new approach for the 21st century. In: Marin GB (ed) Advances in chemical engineering. Academic Press, Dordrecht, pp 1–313

Green WH, Barton PI, Bhattacharjee B et al (2001) Computer construction of detailed chemical kinetic models for gas-phase reactors. Ind Eng Chem Res 40:5362–5370

Grenda JM, Androulakis IP, Dean AM et al (2003) Application of computational kinetic mechanism generation to model the autocatalytic pyrolysis of methane. Ind Eng Chem Res 42:1000–1010

Gueniche HA, Biet J, Glaude PA et al (2009) A comparative study of the formation of aromatics in rich methane flames doped by unsaturated compounds. Fuel 88:1388–1393

Hakka MH, Bennadji H, Biet J et al (2010) Oxidation of methyl and ethyl butanoates. Int J Chem Kinet 42:226–252

Halstead MP, Kirsch LJ, Prothero A et al (1975) A mathematical model for hydrocarbon autoignition at high pressures. Proc Roy Soc A Math 346:515–538

Harper MR, Van Geem KM, Pyl SP et al (2011) Comprehensive reaction mechanism for n-butanol pyrolysis and combustion. Combust Flame 158:16–41

Herbinet O, Biet J, Hakka MH et al (2011) Modeling study of the low-temperature oxidation of large methyl esters from C11 to C19. Proc Combust Inst 33:391–398

Herbinet O, Husson B, Serinyel Z et al (2012) Experimental and modeling investigation of the low-temperature oxidation of n-heptane. Combust Flame 159:3455–3471

Heyberger B, Battin-Leclerc F, Warth et al (2001) Comprehensive mechanism for the gas-phase oxidation of propene. Combust Flame 126:1780–1802

Homer RW, Swanson J, Jilek RJ et al (2008) SYBYL line Notation (SLN): a single notation to represent chemical structures, queries, reactions, and virtual libraries. J Chem Inf Model 48:2294–2307

Hoyermann K, Mauß F, Zeuch T (2004) A detailed chemical reaction mechanism for the oxidation of hydrocarbons and its application to the analysis of benzene formation in fuel-rich premixed laminar acetylene and propene. Phys Chem Chem Phys 6:3824

Hu H, Keck J (1987) Autoignition of adiabatically compressed combustible gas mixtures. SAE International. n° 872110, Warrendale

Husson B, Ferrari M, Herbinet O et al (2013) New experimental evidence and modeling study of the ethylbenzene oxidation. Proc Combust Inst 34:325–333

Jalan A, West RH, Green WH (2013) An extensible framework for capturing solvent effects in computer generated kinetic models. J Phys Chem B 117:2955–2970

Kee RJ, Rupley FM, Miller JA (1993) Chemkin II. A fortran chemical kinetics package for the analysis of a gas-phase chemical kinetics. Sandia Laboratories Report, SAND 89-8009B

Lewis GN (1916) The atom and the molecule. J Am Chem Soc 38:762–785

Magoon GR, Green WH (2013) Design and implementation of a next-generation software interface for on-the-fly quantum and force field calculations in automated reaction mechanism generation. Comput Chem Eng 52:35–45

Matheu DM (2002) Integrated pressure-dependence in automated mechanism generation: a new tool for building gas-phase kinetic models. Ph.D. thesis, Massachusetts Institute of Technology. Department of Chemical Engineering

Matheu DM, Lada TA, Green WH et al (2001) Rate-based screening of pressure-dependent reaction networks. Comput Phys Commun 138:237–249

Matheu DM, Dean AM, Grenda JM et al (2003) Mechanism generation with integrated pressure dependence: a new model for methane pyrolysis. J Phys Chem A 107:8552–8565

Miller JA, Klippenstein SJ (2012) Comment on automatic estimation of pressure-dependent rate coefficients (Allen JW, Goldsmith CF, Green WH Phys Chem Chem Phys 14:1131–1155 (2011)). Phys Chem Chem Phys 14:843–8433

Moreac G, Blurock ES, Mauss F (2006) Automatic generation of a detailed mechanism for the oxidation of n-decane. Combust Sci Technol 178:2025–2038

Morgan HL (1965) The generation of a unique machine description for chemical structures-a technique developed at chemical abstracts service. J Chem Doc 5:107–113

Moss JT, Berkowitz AM, Oehlschlaeger MA et al (2008) An experimental and kinetic modeling study of the oxidation of the four isomers of butanol. J Phys Chem A 112:10843–10855

Muller C, Michel V, Scacchi G et al (1995) A computer program for the evaluation of thermochemical data of molecules and free radicals in the gas phase. J Chim Phys 92:1154–1177

Murray-Rust P, Rzepa HS (1999) Chemical Markup, XML, and the worldwide web. 1. Basic principles. J Chem Inf Comput Sci 39:928–942

Nilsson NJ (1982) Principles of artificial intelligence. Birkhäuser, Basel

Pousse E, Porter R, Warth V et al (2010) Lean methane premixed laminar flames doped by components of diesel fuel II: *n*-propylcyclohexane. Combust Flame 157:75–90

Ranzi E, Faravelli T, Gaffuri P et al (1995a) Low-temperature combustion: automatic generation of primary oxidation reactions and lumping procedures. Combust Flame 102:179–192

Ranzi E, Gaffuri P, Faravelli T et al (1995b) A wide-range modeling study of n-heptane oxidation. Combust Flame 103:91–106

Ranzi E, Frassoldati A, Granata S et al (2005) Wide-range kinetic modeling study of the pyrolysis, partial oxidation, and combustion of heavy *n*-alkanes. Ind Eng Chem Res 44:5170–5183

Ratkiewicz A, Truong TN (2006) Automated mechanism generation: from symbolic calculation to complex chemistry. Int J Quantum Chem 106:244–255

Sarathy SM, Westbrook CK, Mehl M et al (2011) Comprehensive chemical kinetic modeling of the oxidation of 2-methylalkanes from C7 to C20. Combust Flame 158:2338–2357

Sirjean B, Buda F, Hakka H et al (2007) The autoignition of cyclopentane and cyclohexane in a shock tube. Proc Combust Inst 31:277–284

Smith EG, Wiswesser WJ, Addelston A (1968) The Wiswesser line-formula chemical notation. McGraw-Hill, New York

Song J, Sumathi R, Yu J et al (2004) Next generation model construction software, and new approaches to estimating rates and thermochemistry for combustion. Abstr Pap Am Chem Soc 228:U233

Spialter L (1963) The atom connectivity matrix (ACM) and its characteristic polynomial (ACMCP): a new computer-oriented chemical nomenclature. J Am Chem Soc 85:2012–2013

Spialter L (1964) The atom connectivity matrix characteristic polynomial (ACMCP) and its physico-geometeric (topological) significance. J Chem Doc 4:269–274

Stein S, Heller S, Tchekhovski D (2003) An open standard for chemical structure representation—the IUPAC chemical identifier. In: Proceedings of the 9th international chemical information conference, pp 131–143

Susnow RG, Dean AM, Green WH et al (1997) Rate-based construction of kinetic models for complex systems. J Phys Chem A 101:3731–3740

Touchard S, Fournet R, Glaude PA et al (2005) Modeling of the oxidation of large alkenes at low temperature. Proc Combust Inst 30:1073–1081

Troe J (1974) Fall-off curves of unimolecular reaction. Ber Bunsenges Phys Chem 78:478–488

Ugi I, Wochner M (1988) Molecular logic and computer assistance in chemistry. J Mol Struct THEOCHEM 165:229–242

Ullmann JR (1976) An algorithm for subgraph isomorphism. J ACM 23:31–42

Van Geem KM, Reyniers MF, Marin GB et al (2006) Automatic reaction network generation using RMG for steam cracking of n-hexane. AIChE J 52:718–730

Warr WA (2011) Representation of chemical structures. Wiley Interdisc Rev Comput Mol Sci 1:557–579

Warth V, Stef N, Glaude PA et al (1998) Computed aided design of gas-phase oxidation mechanisms: application to the modelling of normal-butane oxidation. Combust Flame 114:81–102

Warth V, Battin-Leclerc F, Fournet R et al (2000) Computer based generation of reaction mechanisms for gas-phase oxidation. Comput Chem 92:1666–1692

Weininger D (1988) SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. J Chem Inf Comput Sci 28:31–36

Weininger D, Weininger A, Weininger JL (1989) SMILES. 2. Algorithm for generation of unique SMILES notation. J Chem Inf Comput Sci 29:97–101

Westbrook C, Pitz W, Herbinet O et al (2009) A comprehensive detailed chemical kinetic reaction mechanism for combustion of n-alkane hydrocarbons from n-octane to n-hexadecane. Combust Flame 156:181–199

Willett P (2008) From chemical documentation to chemoinformatics: 50 years of chemical information science. J Inf Sci 34:477–499

Wipke WT, Howe WJ (eds) (1977) Computer assisted organic synthesis. ACS symposium series 61. American Chemical Society, Washington, DC

Wiswesser WJ (1982) How the WLN began in 1949 and how it might be in 1999. J Chem Inf Comput Sci 22:88–93

Wiswesser WJ (1985) Historic development of chemical notations. J Chem Inf Mod 25:258–263