



AFRL-OSR-VA-TR-2015-0169

---

**Automated discovery of new chemical reactions and accurate calculation of their rates**

**William Green**  
**MASSACHUSETTS INSTITUTE OF TECHNOLOGY**

---

**06/02/2015**  
**Final Report**

DISTRIBUTION A: Distribution approved for public release.

Air Force Research Laboratory  
AF Office Of Scientific Research (AFOSR)/ RTE  
Arlington, Virginia 22203  
Air Force Materiel Command

<b>REPORT DOCUMENTATION PAGE</b>				<i>Form Approved</i> OMB No. 0704-0188	
<p>The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing the burden, to Department of Defense, Executive Services, Directorate (0704-0188). Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.</p> <p><b>PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ORGANIZATION.</b></p>					
<b>1. REPORT DATE (DD-MM-YYYY)</b> 15-07-2015		<b>2. REPORT TYPE</b> Final Performance		<b>3. DATES COVERED (From - To)</b> 01-03-2013 to 28-02-2015	
<b>4. TITLE AND SUBTITLE</b> Automated discovery of new chemical reactions and accurate calculation of their rates				<b>5a. CONTRACT NUMBER</b>	
				<b>5b. GRANT NUMBER</b> FA9550-13-1-0065	
				<b>5c. PROGRAM ELEMENT NUMBER</b>	
<b>6. AUTHOR(S)</b> William Green				<b>5d. PROJECT NUMBER</b>	
				<b>5e. TASK NUMBER</b>	
				<b>5f. WORK UNIT NUMBER</b>	
<b>7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)</b> MASSACHUSETTS INSTITUTE OF TECHNOLOGY 77 MASSACHUSETTS AVE CAMBRIDGE, MA 02139-4301 US				<b>8. PERFORMING ORGANIZATION REPORT NUMBER</b>	
<b>9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)</b> AF Office of Scientific Research 875 N. Randolph St. Room 3112 Arlington, VA 22203				<b>10. SPONSOR/MONITOR'S ACRONYM(S)</b> AFOSR	
				<b>11. SPONSOR/MONITOR'S REPORT NUMBER(S)</b>	
<b>12. DISTRIBUTION/AVAILABILITY STATEMENT</b> A DISTRIBUTION UNLIMITED: PB Public Release					
<b>13. SUPPLEMENTARY NOTES</b>					
<b>14. ABSTRACT</b> A major challenge in combustion chemistry is to be sure that all of the important reactions have been included in the kinetic model. Experience shows that missing reactions are a major cause of discrepancies between model predictions and experimental performance. Looking at combustion chemistry as motion on a Potential Energy Surface landscape, reactions are the passes that connect valleys representing the different molecules; whether or not two valleys are connected is the key issue determining the topology of the landscape. Conventional kinetic modeling approaches include only well known types of reactions in the models; as demonstrated in this project these conventional models omit most of the actual reactions occurring in the engine. Discovering new reactions is very challenging, so historically many of the most important reactions have been discovered by serendipity (either experimentally or by using quantum chemistry).					
<b>15. SUBJECT TERMS</b> combustion chemistry, kinetics, reaction pathways					
<b>16. SECURITY CLASSIFICATION OF:</b>			<b>17. LIMITATION OF ABSTRACT</b>	<b>18. NUMBER OF PAGES</b>	<b>19a. NAME OF RESPONSIBLE PERSON</b> William Green
<b>a. REPORT</b>	<b>b. ABSTRACT</b>	<b>c. THIS PAGE</b>			<b>19b. TELEPHONE NUMBER (Include area code)</b> 617-253-4580
U	U	U	UU		

FA9550-13-1-0065

Program Manager: Dr. Chiping Li

## **Automated discovery of new chemical reactions and accurate calculation of their rates**

William H. Green, Hoyt C. Hottel Professor of Chemical Engineering  
Massachusetts Institute of Technology

### **Final Progress Report**

#### ***Executive Summary***

A major challenge in combustion chemistry is to be sure that all of the important reactions have been included in the kinetic model. Experience shows that missing reactions are a major cause of discrepancies between model predictions and experimental performance. Looking at combustion chemistry as motion on a Potential Energy Surface landscape, reactions are the passes that connect valleys representing the different molecules; whether or not two valleys are connected is the key issue determining the topology of the landscape. Conventional kinetic modeling approaches include only “well known” types of reactions in the models; as demonstrated in this project these conventional models omit most of the actual reactions occurring in the engine. Discovering new reactions is very challenging, so historically many of the most important reactions have been discovered by serendipity (either experimentally or by using quantum chemistry). In this project we develop a much more systematic approach, using high performance computing to search for reactions connecting a specified reactant to each possible product, and using quantum chemistry to determine the reaction barriers. The algorithm developed here has 4 key steps: (1) identification of all the potential product channels; (2) Search for a low energy reaction path connecting reactant and each product channel; and (3) refinement of the reaction path to determine the saddle point geometry and barrier height. This new computational approach is found to be the most effective method so far at discovering novel chemical reactions. Using the new algorithm we have discovered 38 new exothermic or mildly endothermic reactions in this project; about half of these newly-discovered reactions were completely surprising. Possible future research directions to make this promising algorithm even more reliable, computationally efficient, and useful for engine modeling purposes are discussed briefly.

## Introduction

Currently, new chemical understanding is driven mainly by the experimental discovery of new compounds and their reactivity. This experimental work is expensive and time-consuming, but the payoff for discovering new chemical reactions is enormous in terms of both fundamental understanding and practical engineering. Examples of important new reactions discovered in recent decades include many advances that followed from the discoveries of olefin metathesis<sup>1</sup> and click chemistry.<sup>2</sup> Theoretical, algorithmic and computational advances in quantum chemistry over the last two decades have brought it to the stage where it is widely applied for qualitative interpretation of such new experimental findings, and increasingly used to construct detailed quantitative descriptions of chemical reaction networks.

The main parameters in chemical kinetic models – equilibrium constants and rate coefficients – are increasingly derived from quantum mechanical calculations on stationary points (local minima and first-order saddle points) along the potential energy surface (PES). Local minima, i.e. reactants, intermediates, and products, are generally easy to characterize due to simple bonding and because the negative of the gradient along the PES always points downhill. Several robust algorithms were developed for identifying local minima.<sup>3</sup> Saddle points, which correspond to transition states, remain a challenge for systematic characterization due to non-standard variations in reactive centers and because a transition structure optimization must step uphill in one direction and downhill in all other orthogonal directions. At present, most saddle points are found based on human intuition, derived primarily from experience on analogous reactions: the human provides a very good initial guess at the saddle point geometry, and then various algorithms (so-called single-ended) are used to refine the initial guess geometry to find the true saddle point.<sup>4-8</sup> This has typically been applied to cases where humans expect a reaction to connect the reactant(s) and product(s), and so it also rarely discovers anything surprising. This conventional approach turns out to be quite good for finding “expected reaction” saddle points, but rarely discover anything that is unexpected. As an alternative to human input, the initial guess structure for the saddle point can be generated by methods (so-called double-ended searches) which reconstruct the reaction path between known reactant and product sides.<sup>9-13</sup>

Most large chemical kinetic models are constructed in ways which do not allow any unexpected reactions to be included; this is true whether the model is constructed by hand or using automatic mechanism generators.<sup>14</sup> Incorporation of even a simple new reaction in chemical kinetic models can significantly alter these models since almost certainly it will be the first member of a new large family of related reactions. For instance, several unexpected low-barrier reactions for JP-10 pyrolysis<sup>15</sup> and for decomposition of the ketohydroperoxides,<sup>16</sup> which are critical intermediates in alkane autooxidation and ignition chemistry, were recently discovered in our group. Including the new unexpected reactions and analogous new reactions in the kinetic models significantly improved the reliability of the kinetic model predictions.<sup>17</sup>

All this experience suggests that if we search for them systematically, we can expect to find many “unexpected” reactions which are omitted from existing chemistry models, and that incorporating these new reactions will significantly improve the reliability of the model predictions. In the present work, we propose a systematic approach for discovery new chemical reactions with minimum human effort. The method advances recent work on saddle point optimization algorithms, by making use of both single- and double-ended methods in a fully automated procedure. Among a great variety of methods,<sup>4-13</sup> we employ a combination of the Broyden optimization algorithm as implemented and improved by Schlegel<sup>6-8</sup> and the freezing string method<sup>13</sup> as both methods were readily available to us at the time we started this project. We find that this combination works fairly well for several systems of combustion and atmospheric chemistry importance. We were able to detect not all ‘known’ reaction pathways, manually detected in the previous

studies, but also new, previously 'unknown', reaction pathways, which involve significant atom rearrangements. The present work is complementary to recent developments in computational chemistry<sup>14,18</sup> which demonstrate a tendency of switching from a traditional secondary role of theory in interpreting experimental findings in chemistry to a powerful predictive tool for analysis of chemical reactivity, discovery of chemical reaction pathways, and generation of reaction mechanisms prior to any experiments.

## Computational Approach

### *Overview*

We have developed algorithms for comprehensively enumerating possible products from a given reactant using Ugi's matrix representation of both molecular structures and chemical reactions. The products are sorted by thermochemistry (so the most exothermic reactions are explored first; by the Hammond postulate these usually have the lowest barriers) and by the number of bonds broken and formed (which correlates with the Arrhenius A factor). The thermochemistry is computed using Benson-type group additivity methods or quantum chemistry at the user's option. The computer then automatically searches for a low-energy reaction path connecting the reactant to the products, using the Freezing-String Algorithm and Density Functional Theory. The reaction path is then refined using a Berny-type search to find the saddle point (i.e. the top of the mountain pass), and the barrier height is computed using the user-specified level of quantum chemistry.

The overall algorithm was designed and implemented by this project. To speed the development and to leverage prior research, we connected our software with several different packages: RMG for estimating thermochemistry, MarvinBeans for converting the matrix representations into initial-guess geometries for each conformation, the Merck force field MMFF94 for refining the initial geometries, Q-Chem for the Freezing-String reaction path search using DFT, and GAUSSIAN for refining the transition state and for high-accuracy (e.g. CCSD(T)) barrier height calculations.

### *Molecules Used to Test the New Algorithm*

We have applied the new algorithm to reactions of several different types of chemical species expected to be important in engines: (1) the ketohydroperoxide  $\text{HOOCH}_2\text{CH}_2\text{CHO}$ , representative of the chain-branching species important in diesel ignition; (2) N-(hydroperoxymethyl) formamide  $\text{HOOCH}_2\text{NHCHO}$ , representative of the key intermediate in oxidative degradation of fuel detergents and the dispersants in jet and diesel engine lubricants; (3) the cyclic peroxides formed from  $\text{HOOCH}_2\text{CH}_2\text{CHO}$  and  $\text{HOOCH}_2\text{NHCHO}$ ; (4) the related secondary ozonide and the corresponding N-containing compound formed in pollutant plumes; (5) 1,4-pentadiene, representative of multiply unsaturated hydrocarbons formed in fuel-rich combustion. These test molecules were chosen to cover a wide range of chemistry important in engines, and a variety of molecular structures, including heteroatoms and a mix of cyclic and acyclic species.

### *Technical Details of the Computations*

## 1. Matrix Representation of Species and Reactions

There are different ways to describe mathematically compounds presented in a system.<sup>19-26</sup> In the present algorithm, we chose the Bond Electron (BE) matrices<sup>19,20</sup> as one of the most vivid and convenient representations. A BE matrix  $\mathbf{R}$  for a compound with  $n$  atoms  $r_1, r_2, \dots, r_n$  is a square matrix of dimension  $n$ . The  $i^{\text{th}}$  row of this matrix corresponds to electron configuration of atom  $r_i$ . The  $R_{ij}$  element represents the number of covalent bonds between atoms  $r_i$  and  $r_j$  and matrix  $\mathbf{R}$  is symmetric, i.e.  $R_{ij} = R_{ji}$ . The diagonal elements  $R_{ii}$  represent the number of free valence electrons which are not involved in bonds. The sum over the entries of the  $i^{\text{th}}$  row or column of a BE-matrix is the number of valence electrons which belong to the atom  $r_i$ .

The total matrix of the initial reactive system  $\mathbf{R}_{\text{sum}}$  is depicted by a block matrix, each block representing one compound. A chemical reaction is the conversion of the system by the redistribution of valence electrons. It is represented by transformation of the initial matrix  $\mathbf{R}_{\text{sum}}$  by the addition of the reaction matrix  $\mathbf{A}$  to yield a product total matrix  $\mathbf{P}_{\text{sum}}$  with altered connectivity. The reaction matrix  $\mathbf{A}$  is also symmetric and its elements represent the appearance (positive value) or the disappearance (negative value) of the localized valence electrons. This matrix should conserve the total number of electrons in the system, i.e.

$$\sum_{i,j} a_{ij} = 0$$

Usually in elementary reactions  $|a_{ij}| < 3$  since the change by two implies a significant chemical process, for instance, formation or rupture of a double bond in a single elementary step.

After applying the reaction matrix  $\mathbf{A}$ , the product matrix  $\mathbf{P}$  can then be rearranged to form a block structure to represent the product molecules. If the product matrix satisfies the mathematical and chemical constraints, the reaction pathway is considered to be feasible and is added to the list of channels for further thermochemistry calculations.

## 2. Reducing the Size of the Search Space

*Thermochemistry Calculations.* Even for a small hydrocarbon system, the number of reaction pathways which can be generated using the BE matrix representation described above can be massive, increasing exponentially with the number of carbon atoms. The thermodynamic reaction control principle can be implemented to narrow the search space: highly endothermic reactions are likely to be too slow to be important. The relative stability of the products can be estimated using their thermochemical properties, and reaction channels with less stable products can be eliminated. However, on-the-fly quantum-chemical calculation of thermochemical properties for hundreds (or even thousands) species using, e.g. density functional theory (DFT) methods, can also be computationally very demanding. One efficient alternative way to do so is to use group contribution methods that estimate the thermochemistry of a molecule based on the sub-molecular fragments present in the molecule. The Benson group additivity framework is such an example of a group contribution method that has proven to provide accurate estimates of the ideal gas thermochemistry for a large range of molecules. Benson's group additivity approach<sup>27</sup> divides a molecule into functional groups, and the contribution of each functional group to the overall thermochemistry is included. The group additivity enthalpy estimate is typically within a few kcal/mol<sup>28</sup> of the truth, sufficient to reliably separate very endothermic reactions from the rest.

*Generating 3D Geometries.* While DFT methods can, in principle be implemented to generate and optimize the starting and final 3D geometries, the large number of reaction channels can prohibit such calculations. A much faster route is to use a molecular mechanics force field. Though, in principle, any type of reasonable

force field can be used, we choose Merck Molecular force-fields (MMFF94) developed by Merck Research Laboratories<sup>29-33</sup> which were specifically optimized for a wide range of organic chemistry calculations (from very small molecules to proteins). The method includes parameters from high-quality quantum calculations (rather than experimental data) for a wide range of atom types including those presented in the examples studied in this work.

### 3. Locating Transition States

As mentioned in Introduction, methods for transition state structure search can be classified as single-ended and double-ended.<sup>4,5</sup> The single-ended methods refine an initial guess of the transition state to the exact answer by exploring the PES using local gradient and usually the second derivative information. The Bery saddle point optimization algorithm is one of these methods, which is widely used in *ab initio* calculations primarily due its reliability.<sup>6-8</sup> Bery geometry optimization starts with an initial guess for the second derivative matrix (Hessian) which is determined using connectivity derived from atomic radii and a simple valence force field.<sup>7,8</sup> The approximate Hessian matrix is improved at each point using the first derivatives and energies computed along the optimization pathway. The main input for such algorithm is a guess structure that should be located very close to the transition state in order to succeed. At present, initial guess geometries are usually found based on human intuition, and automated generation of good initial guess transition state structures is the main point of the procedure proposed here.

The double-ended methods start from the reactants and products and work from both sides to find the transition structure and the reaction path. These include the nudged elastic band method,<sup>9,10</sup> string method,<sup>11,12</sup> and freezing string method (FSM).<sup>13</sup> All these algorithms generate a sequence of configurations located between the reactant and products geometries. In the present algorithm we choose FSM as it does not require Hessian calculations and provides one of the most inexpensive algorithms for determining a path with a specified number of intermediate structures (nodes) connecting the known reactant(s) and product(s). FSM starts from both ends of the path, and adds nodes irreversibly until the two ends join. FSM partially optimizes intermediate structures orthogonally to the reaction path, and then freezes them before a new pair of points along the reaction path is added. The number of gradients required to locate a guess saddle point using FSM is rather small.

Since the reactant and product geometries, corresponding to initial and final minima, are known from the matrix representation, energy cutoff and force field optimization described above, automatic path-finding tools, such as the freezing string method, can characterize a reaction coordinate joining the end-points. The highest point on the pathway becomes a good initial guess for subsequent refinement of the transition structure using the single-ended method such as Bery algorithm.

#### ***Summary of methodology and computational details***

All the methods just described can be combined in a systematic and automated procedure for identifying reaction pathways. It is summarized in the flowchart presented on Figure 1. Only initial reactant(s) geometry is required at the start. It is converted to the reactant matrix **R**. Different reaction matrices **A** are then generated, which break up to  $N$  bonds and make up to  $M$  bonds, where  $N$  and  $M$  are defined by the user. The number of possibilities rises combinatorially with  $N$  and  $M$ . In the benchmark examples below, we constrain  $1 < N < 4$  and  $1 < M < 4$ . We constrain  $N, M < 4$  to reduce computational demands and  $N, M > 1$  since most of those simple reactions are well understood. To reduce the computational effort we restricted the products to those which can be drawn with a Lewis structure with a fixed number of valence electrons for each element – 4 for carbon, 2 plus two lone pairs for oxygen, 1 for hydrogen, 3 plus a lone pair for nitrogen.

In the present work, we restrict all the elements  $R_{ij}$  and  $P_{ij}$  in the reactant and product matrices to be integers. This can potentially lead to issues for products with resonant structures. We also demand number of unpaired electrons on an atom not exceed 1. This is certainly ruling out some known species, e.g. biradicals such as methylene and  $O(^1P)$ , but we can always relax this restriction and include them in future studies. Since BE matrices are invariant with certain permutations, several BE matrices can correspond to chemically equivalent compounds. However, in order not to omit important reaction pathways, all combinations are taken into account initially. We remove any duplicates before the molecular mechanics and quantum chemistry calculations are run.

The product matrices  $\mathbf{P}$  are obtained and converted to block structure by simple linear algebra operations. Thermochemistry calculations are used to narrow the search space - during this step 'known' and unstable products are detected. In the present study, a reaction is considered to be too endothermic to be interesting if the standard enthalpy of reaction (denoted as  $\Delta H^\ddagger$  in Tables 2-5) is higher than 20 kcal/mol. These calculations are performed using the Reaction Mechanism Generator (RMG) software package developed in our group.<sup>14</sup>

The next step is automated saddle point search for 'unknown' reaction pathways which lead to stable products. For quantum chemistry calculations, geometries of reactant(s) and product(s) are required. They are generated by converting corresponding blocks of the  $\mathbf{P}$  matrices to SMILES strings (standard format for molecule mechanics calculations) and quick force field optimization using academic version of the MarvinBeans package from ChemAxon.<sup>34</sup> Special care is taken of the conversion between different molecule representation formats, eliminating duplicate channels. In the present study, we employ the MMFF94 force field to determine the geometry of the products.

The BE matrix representation is identical for all conformers. However, the distinct 3d structure of conformers strongly affects the properties and the reactions of molecules. In order to take into account conformational effects, we used the MarvinBean's Conformer Plugin to generate a series of stable 3d structures (up to three in the present study), i.e. conformers for each compound. If more than one product is formed in the reaction channel, their structures were first aligned to maximum coincidence with the reactant structure in non-mass weighted Cartesian coordinates and then the length of the vector between the centers of mass of the products was increased in order to avoid overlaps between atoms in different product molecules.

For the determination of all the transition states we used the FSM implemented in the Q-Chem software package with default spacing parameters (20 nodes, 6 perpendicular gradients per node).<sup>35</sup> In order not to repeat computations on duplicate structures, all input structures are compared prior to initiating FSM calculations and identical input files were eliminated. However, chemically equivalent channels with distinct 3d geometries were not excluded from the procedure in order not to omit conformers or atom arrangements which could potentially lead to low-energy reaction pathways. For channels where FSM found an apparent single barrier along the reaction path (without any restrictions to the energy barrier height), these calculations were followed by transition state search using the Berny optimization algorithm implemented in the Gaussian 03 software package<sup>36</sup> starting from the FSM geometry. For each detected saddle point, we also perform intrinsic reaction path (IRC) calculations which go downhill from the transition state to see which two minima it is connected to and to verify whether this saddle point corresponds to the reaction from  $R$  to  $P$ .

Geometry optimization and reaction path analysis for the examples herein were performed using M062X/6-311++G\* level of theory, while single point energies at the stationary points were obtained using CCSD(T)/6-311++G\* level of theory. Thermochemistry of the detected reactions was computed using CBS-QB3 level of theory. Note that the proposed procedure is not dependent on the use of any particular software package or

level of theory. Our choice of such parameters as quantum chemistry methods, basis set, force fields, FSM parameters, number of conformers, method for orientating product structures should be considered as reasonable first attempts which can be improved in the future as the community gains more experience.

Each set of quantum chemistry calculations – structure optimization, energy calculation and the FSM calculation – is performed independently in parallel to exploit the full potential of supercomputers. All the computations are automated, to reduce time-to-completion and also to reduce bias related to manual intervention to help the computer find a saddle point, and which so tend to find only “human-expected” known reactions. In order to demonstrate how such fully-automated procedure can accelerate the discovery of new chemistry, we use it to study several systems of combustion, oxidation and atmospheric chemistry importance which span a reasonable wide space of chemical reactivity.

## Results

For each initial structure, Table 1 shows the number of product channels generated using the BE representation. The results of thermochemistry analysis using Benson group additivity approach of all product channels are also given in Table 1. All reaction pathways identified are summarized in Figures 2-5 and Tables 2-5.

### 1. New reactions of Ketohydroperoxides

Ketohydroperoxides are the main source of radicals during the first stage of low-temperature ignition, and are also formed in liquid phase oxidation. Combustion parameters, such as ignition delay, are very sensitive to the details of their chemistry.<sup>16</sup> Most previous combustion models assumed ketohydroperoxides do only one reaction: forming radicals by breaking the weak O-O bond. However, new chemical reactions of ketohydroperoxides which transform them into cyclic peroxides and then acids and carbonyl molecules has been recently characterized<sup>16</sup> suggesting that chemistry of ketohydroperoxides is more complicated.

Table 1 shows that almost 500 product channels, which break and make just 2 or 3 bonds, are possible from this small molecule. (Note that this includes distinct reactions which form the same products, i.e. there is often more than one mountain pass connecting two valleys. Without repeated channels, this number reduces to 109.). Half of the channels are exothermic and almost 70 % of the channels have  $\Delta H_{rxn} < 20$  kcal/mol. At the same time, only one (!) product channel (1,2-dioxolan-3-ol) is reached by a “well-known” reaction type available in the current RMG database – the first step of the Korcek mechanism forming an intermediate cyclic peroxide.<sup>16</sup> As shown in Figure 2 (Reaction (1)), our simple algorithm was able to locate efficiently transition states for this ‘known’ reaction though it involves a significant rearrangement of atoms including cycle formation and hydrogen transfer between two oxygen atoms. Figure 2 also shows that the automatic search found transition states for 5 additional reaction pathways. Note that one of these channels leads directly to the products of the Korcek mechanism - Reaction (4) on Figure 2. However, Table 2 shows that the energy barrier (denoted as  $\Delta E^\ddagger$ ) for this channel is higher by almost 20 kcal and therefore, rather than going directly over this high mountain pass, sequential reactions via an intermediate valley (the cyclic peroxide) are expected to be the fastest reaction path. This channel was located while searching for a reaction path to a different product (see Comments in Table 2). This example illustrates that the saddle-point search procedure used in this work does not guarantee convergence to the desired transition state. As a result of its non-

iterative nature, the FSM cannot guarantee conservation of the exact IRC profile and therefore further optimization of the guess structure (the highest energy node) with Berny optimization can lead to structures which do not correspond to the initial reaction path. The calculations also occasionally converged to saddle points which correspond to different initial reactants. More importantly, the majority of our reaction path searches do not converge to a saddle point. This typically occurred because the energy profiles from the FSM calculations for these paths exhibited several very high barriers – not the single elementary-step transition state we were looking for – and therefore the Berny-type TS search was not performed. In these cases, it is possible that no single saddle point exists that connects reactant and product(s), or just that with our choices of parameters and options the FSM is not robust enough to find one. Clearly improved reaction path search methods are needed. Nonetheless the current algorithm is sufficient to discover several new unexpected chemical reactions.

## 2. Cyclic peroxides and ozonides

Upon examination of the reactivity of 1,2-dioxolan-3-ol – our second case study - we confirmed that fragmentation of the cyclic peroxide leads to two possible pairs of acid and aldehyde products (Reactions (24) and (25) in Figure 4) previously characterized in Ref. 16. We were also able to locate the reverse pathway which leads to the initial molecule (Reaction (22) in Figure 4) confirming that the proposed procedure is rather sustainable. The fact that the fully automated search reproduces the results of the previous manual studies of the Korcek mechanism – and also discovers a half-dozen additional transition states to products not found by that manual search – is encouraging.

Another interesting example is the chemistry of the Criegee intermediate with formaldehyde (third case study). These two species quickly recombine forming the more stable secondary ozonide intermediate (ethylene ozonide)[38] whose reactivity was investigated in the present work. All three reaction pathways identified for ethylene ozonide are presented in Figure 3 (Reactions (15)-(17)) and Table 3. In accordance with the previous manual search [38], we find that two low energy pathways for ethylene ozonide decomposition are formation of hydroxymethyl formate by breaking the weak O-O bond (the lowest energy saddle point) and formation of formaldehyde and formic acid. In addition to the reactions found in Ref. 38, we also find that ethylene ozonide can decompose to formaldehyde and oxirane with similar barrier height.

## 3. Nitrogen-containing Analogues

Table 1 shows that the chemistry of the N-containing analogous compounds is even less known – no unimolecular channels with 2 and 3 bonds breaking/forming were found in the RMG database. However, we were able to identify and characterize several reaction pathways using our automated procedure. The results are summarized in Tables 2-4 and Figures 2-4. They show that the chemistry for these molecules is similar to the first three structures. For example, Reaction (7) is similar to the first step of the Korcek mechanism (Reaction (1)). However, replacing one CH<sub>2</sub> group with an NH group significantly enriches the chemistry. For example, Reaction (9) involves two H-transfers (transition state shown in Figure 2) and Reactions (11) and (12) form 3 products from one reactant.

Table 2 shows that in the case of N-(hydroperoxymethyl) formamide, half of the identified saddle points do not correspond to the intended product(s) generated by the BE matrix representation. Due to initial open-chain structure, all these cyclization reactions require significant rearrangement of several atoms, and apparently FSM has trouble with these cases. Similarly, the algorithm failed to find several cyclization

transition states of the  $\gamma$ -ketohydroperoxide. At the same time, all saddle points detected for more constrained cyclic reactants 1,2,4-dioxazolidine and 1,2,4-dioxazolidin-3-ol are for the intended products (see Tables 3 and 4).

#### 4. 1,4-Pentadiene

As a final example, the chemical reactivity of 1,4-pentadiene is investigated (see Table 5 and Figure 5). Due to many possible single and double hydrogen atom transfers, many product channels are possible in this system. Table 1 shows that total number of channels exceeds seven hundred. However, out of this number, only 8 saddle points were located and 3 of these correspond to known reaction types in the RMG database. All structures were connected to the initial molecule by a large barrier (higher than 60 kcal/mol). Nevertheless, the ability of our algorithm not only to reproduce the existing 'known' reaction pathways but also to identify 5 new reaction pathways for this simple molecule is very promising.

## Conclusions

The algorithm proposed here for finding elementary reaction paths possesses several important features: it requires no human intervention and it does not require any information on the chemical reactivity of the reactant(s). In order to detect saddle points it generates a series of product(s) and tries to find pathways that connect the product(s) to the initial structure - this would be tedious to do without automation. It is much more computationally efficient (i.e. requires fewer energy/gradient calculations) than methods based on *ab initio* molecular dynamics, since it rapidly finds and focuses on the structures near the saddle point connecting the reactant with designated products. The present examples show that our automated algorithm is able to detect not only previously detected reaction pathways but also several new, previously unknown, types of reactions. Our constraint that allows only two or three bonds to be formed and broken in the same elementary step worked well for the present examples, keeping the number of possible product channels manageable (< 1000 possible reaction paths per reactant) but still allowing us to discover a lot of new chemistry. The present results also show that detecting thermodynamically feasible products does not guarantee kinetic feasibility. The degree of success in locating unexpected kinetically favorable elementary steps is rather low – a significant number of thermodynamically stable structures either are connected to the initial structure only through multiple elementary steps or via chemically infeasible routes. In future, some of these routes which are the most obviously unlikely, such as breaking the C=O bond, might be eliminated in order to focus resources on discovering transition states for more likely reactions. Another possibility is to freeze certain bonds and angles that are expected to be unreactive in order to reduce the number of conformers and matrices dimensionality. However, such limitations should be imposed with caution as they may prohibit detecting unexpected important reaction pathways!

The success of the exact search for saddle point relies on the ability to generate very accurate corresponding guess structures from a double-ended method, such as the freezing string method (FSM) employed in the present study. FSM can be considered as one of the fastest double-ended methods. However, FSM is not guaranteed to converge to the reaction path of interest and therefore convergence to the correct saddle point cannot be rigorously guaranteed.<sup>13</sup> For each detected transition state we performed the intrinsic reaction coordinate (IRC) integration calculation to recover both reactant and product sides. In several cases, we found that the Bery optimization of the highest node along the reaction path generated by FSM leads to wrong transition states, which do not connect the specified reactant and product states. Therefore, a good initial

guess of the transition state remains one of the major challenges in the use of such algorithms for complex systems.<sup>39</sup>

As an alternative to FSM, its predecessor, the growing string method (GSM), an iterative algorithm that connect the reactant and the product via the IRC, can be implemented. Recent advances in the growing string method have allowed it to locate the exact transition state for a given reaction path.<sup>39,40</sup> A procedure for automatic reaction path finding which is based on GSM has been recently proposed.<sup>41</sup> However, while formally very attractive, the GSM is substantially more computationally expensive compared to the cost of FSM calculation (by up to a factor of 10), and in this work the FSM calculations are the limiting computational step in the algorithm. Moreover, even GSM can fail to locate a saddle point for a single elementary step as no double-ended string method is perfectly reliable. An important point of the present testing is that FSM is able to discover several important transition states recently discovered by manual searches.<sup>16,38</sup> However, future improvement of reliability as well as reducing computational costs of double-ended methods is obviously required.

## References

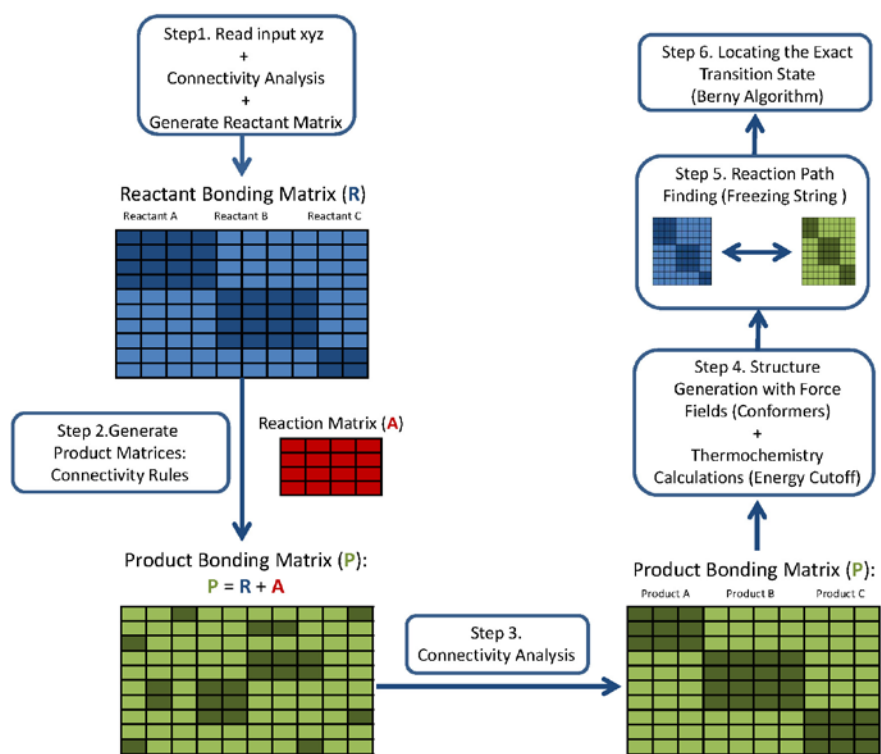
- [1] Schrodri, Y.; Pederson, R. L. Evolution and Applications of Second-Generation Ruthenium Olefin Metathesis Catalysts. *Aldrichimica Acta* **2007**, *40*, 45-52.
- [2] Kolb, H. C.; Finn, M. G.; Sharpless, K. B. Click Chemistry: Diverse Chemical Function from a Few Good Reactions. *Angew. Chem. Int. Ed.* **2001**, *40*, 2004–2021.
- [3] Magoon, G. R.; Green, W. H. 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.* **2013**, *52*, 35-45.
- [4] Wales, D. J. Energy landscapes: calculating pathways and rates. *Int. Rev. Phys. Chem.* **2006**, *25*, 237-282
- [5] Wales, D. J. Energy Landscapes. *Cambridge University Press*, Cambridge, **2003**.
- [6] Schlegel, H. B. Optimization of Equilibrium Geometries and Transition Structures. *J. Comp. Chem.* **1982**, *3*, 214-218.
- [7] Peng, C.; Ayala, P. Y.; Schlegel, H. B.; Frisch, M. J. Using redundant internal coordinates to optimize equilibrium geometries and transition states. *J. Comp. Chem.* **1996**, *17*, 49-56.
- [8] Schlegel, H. B. Estimating the Hessian for gradient-type geometry optimizations. *Theor. Chem. Acc.* **1984**, *66*, 333-340.
- [9] Henkelman, G.; Uberuaga, B.P.; Jonsson, H. A climbing image nudged elastic band method for finding saddle points and minimum energy paths. *J. Chem. Phys.* **2000**, *113*, 9901-9904.
- [10] Henkelman, G.; Jonsson, H. Improved tangent estimate in the nudged elastic band method for finding minimum energy paths and saddle points. *J. Chem. Phys.* **2000**, *113*, 9978-9985.
- [11] E, W. N.; Ren, W. Q.; Vanden-Eijnden, E. String method for the study of rare events. *Phys. Rev. B* **2002**, *66*, 052301.
- [12] Sheppard, D.; Terrell, R.; Henkelman, G. Optimization methods for finding minimum energy paths. *J. Chem. Phys.* **2008**, *128*, 134106.
- [13] Behn, A.; Zimmerman, P. M.; Bell, A. T.; Head-Gordon, M. Efficient exploration of reaction paths via a freezing string method. *J. Chem. Phys.* **2011**, *135*, 224108.

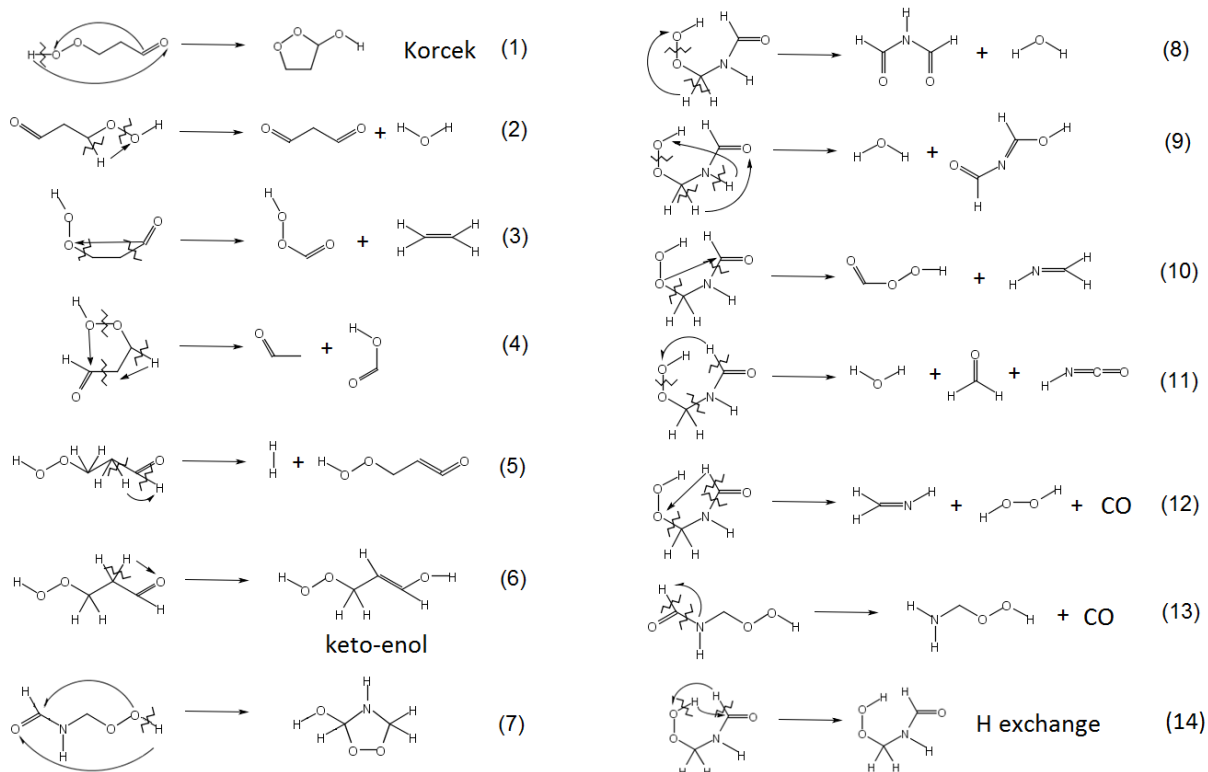
- [14] Green, W. H.; Allen, J. W.; Buesser, B. A.; Ashcraft, R. W.; Beran, G. J.; Class, C. A.; Gao, C.; Goldsmith, C. F.; Harper, M. R.; Jalan, A.; Keceli, M.; Magoon, G. R.; Matheu, D. M.; Merchant, S. S.; Mo, J. D.; Petway, S.; Raman, S.; Sharma, S.; Song, J.; Suleimanov, Y. V.; Van Geem, K. M.; Vandeputte, A. G.; Wen, J.; West, R. H.; Wong, A.; Wong, H.-W.; Yelvington, P. E.; Yee, N.; Yu, J. RMG – Reaction Mechanism Generator. <https://github.com/GreenGroup/RMG-Py> **2014**.
- [15] Magoon, G.; Aguilera-Iparraguirre, J.; Green, W. H.; Lutz, J. J.; Piecuch, P.; Wong, H.-W.; Oluwole, O. O. Detailed Chemical Kinetic Modeling of JP-10 (exo-tetrahydrodicyclopentadiene) High Temperature Oxidation: Exploring the Role of Biradical Species in Initial Decomposition Steps. *Int. J. Chem. Kinet.* **2012**, *44*, 179-193.
- [16] Jalan, A.; Alecu, I. M.; Meana-Pañeda, R.; Aguilera-Iparraguirre, J.; Yang, K. R.; Merchant, S. S.; Truhlar, D. G.; Green, W. H. New pathways for formation of acids and carbonyl products in low-temperature oxidation: The Korcek decomposition of  $\gamma$ -keto hydroperoxides. *J. Am. Chem. Soc.* **2013**, *135*, 11100-11114.
- [17] Vandewiele, N. M.; Magoon, G. R.; Van Geem, K. M.; Reyniers, M.-F.; Green, W. H.; Marin, G. B. Kinetic modeling of jet propellant-10 pyrolysis. *Energy Fuels* **2015**, *29*, 413-427.
- [18] Wang, L.-P.; Titov, A.; McGibbon, R.; Liu, F.; Pande, V. S.; Martinez, T. J. Discovering chemistry with an ab initio nanoreactor. *Nature Chemistry* **2014**, *6*, 1044-1048.
- [19] Tarjan, R. E. Graph Algorithms in Chemical Computation. Algorithms for Chemical Computations. Christoffersen, R. E., Ed.; American Chemical Society: Washington, DC, **1977**.
- [20] Ugi, I.; Bauer, J.; Brandt, J.; Friedrich, J.; Gasteiger, J.; Jochum, C.; Schubert, W. New Applications of Computers in Chemistry. *Angew. Chem. Int. Ed.* **1979**, *18*, 111-123.
- [21] Yoneda, Y. A Computer Program Package for the Analysis, Creation, and Estimation of Generalized Reactions – GRACE I. Generation of Elementary Reaction Network in Radical Reactions – A/GRACE (I). *Bull. Chem. Soc. Jpn.* **1979**, *52*, 8-14.
- [22] Blurock, E. S. Reaction: System for Modeling Chemical Reactions. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 607-616.
- [23] Quann, R. J.; Jaffe, S. B. Structure-Oriented Lumping: Describing the Chemistry of Complex Hydrocarbon Mixtures. *Ind. Eng. Chem. Res.* **1992**, *31*, 2483-2497.
- [24] Broadbelt, L. J.; Stark, S. M.; Klein, M. T. Computer Generated Reaction Modelling Decomposition and Encoding Algorithms for Determining Species Uniqueness. *Comput. Chem. Eng.* **1996**, *20*, 113-129.
- [25] Warth, V.; Battin-Leclerc, F.; Fournet, R.; Glaude, P. A.; Come, G. M.; Scacchi, G. Computer Based Generation of Reaction Mechanisms for Gas-Phase Oxidation. *Comput. Chem.* **2000**, *24*, 541-560.
- [26] Rangarajan, S.; Bhan, A.; Daoutidis, P. Rule-Based generation of Thermochemical Routes to Biomass Conversion. *Ind. Eng. Chem. Res.* **2010**, *49*, 10459- 10470.
- [27] Benson, S. W. Thermochemical kinetics: methods for the estimation of thermochemical data and rate parameters. 2nd ed. 320 p, New York, Wiley, **1976**.
- [28] Sumathi, R.; Carstensen, H.-H.; Green, W. H. Reaction Rate Prediction via Group Additivity, Part 2: H-Abstraction from Alkenes, Alkynes, Alcohols, Aldehydes, and Acids by H Atoms. *J. Phys. Chem. A* **2001**, *105*, 8969-8984.
- [29] Halgren, T. A. Merck Molecular Force Field. I. Basis, Form, Scope, Parameterization, and Performance of MMFF94. *J. Comput. Chem.* **1996**, *17*, 490-519.
- [30] Halgren, T. A. Merck molecular force field. II. MMFF94 van der Waals and electrostatic parameters for intermolecular interactions. *J. Comput. Chem.* **1996**, *17*, 520-552.

- [31] Halgren, T. A. Merck molecular force field. III. Molecular geometries and vibrational frequencies for MMFF94. *J. Comput. Chem.* **1996**, *17*, 553-586.
- [32] Halgren, T. A.; Nachbar, R. B. Merck molecular force field. IV. conformational energies and geometries for MMFF94. *J. Comput. Chem.* **1996**, *17*, 587-615.
- [33] Halgren, T. A. Merck molecular force field. V. Extension of MMFF94 using experimental data, additional computational data, and empirical rules. *J. Comput. Chem.* **1996**, *17*, 616-641.
- [34] "Marvin was used for drawing, displaying and characterizing chemical structures, substructures and reactions, Marvin 6.2.0, 2014, ChemAxon (<http://www.chemaxon.com>)".
- [35] Shao, Y.; Gan, Z.; Epifanovsky, E.; Gilbert, A. T.B.; Wormit, M.; Kussmann, J.; Lange, A. W.; Behn, A.; Deng, J.; Feng, X.; Ghosh, D.; Goldey, M.; Horn, P. R.; Jacobson, L. D.; Kaliman, I.; Khaliullin, R. Z.; Kuś, T.; Landau, A.; Liu, J.; Proynov, E. I.; Rhee, Y. M.; Richard, R. M.; Rohrdanz, M. A.; Steele, R. P.; Sundstrom, E. J.; Woodcock III, H. L.; Zimmerman, P. M.; Zuev, D.; Albrecht, B.; Alguire, E.; Austin, B.; Beran, G. J. O.; Bernard, Y. A.; Berquist, E.; Brandhorst, K.; Bravaya, K. B.; Brown, S. T.; Casanova, D.; Chang, C.-M.; Chen, Y.; Chien, S. H.; Closser, K. D.; Crittenden, D. L.; Diedenhofen, M.; DiStasio Jr., R. A.; Do, H.; Dutoi, A. D.; Edgar, R. G.; Fatehi, S.; Fusti-Molnar, L.; Ghysels, A.; Golubeva-Zadorozhnaya, A.; Gomes, J.; Hanson-Heine, M. W. D.; Harbach, P. H. P.; Hauser, A. W.; Hohenstein, E. G.; Holden, Z. C.; Jagau, T.-C.; Ji, H.; Kaduk, B.; Khistyayev, K.; Kim, J.; Kim, J.; King, R. A.; Klunzinger, P.; Kosenkov, D.; Kowalczyk, T.; Krauter, C. M.; Lao, K. U.; Laurent, A. D.; Lawler, K. V.; Levchenko, S. V.; Lin, C. Y.; Liu, F.; Livshits, E.; Lochan, R. C.; Luenser, A.; Manohar, P.; Manzer, S. F.; Mao, S.-P.; Mardirossian, N.; Marenich, A. V.; Maurer, S. A.; Mayhall, N. J.; Neuscammann, E.; Melania Oana, C.; Olivares-Amaya, R.; O'Neill, D. P.; Parkhill, J. A.; Perrine, T. M.; Peverati, R.; Prociuk, A.; Rehn, D. R.; Rosta, E.; Russ, N. J.; Sharada, S. M.; Sharma, S.; Small, D. W.; Sodt, A.; Stein, T.; Stück, D.; Su, Y.-C.; Thom, A. J. W.; Tsuchimochi, T.; Vanovschi, V.; Vogt, L.; Vydrov, O.; Wang, T.; Watson, M. A.; Wenzel, J.; White, A.; Williams, C. F.; Yang, J.; Yeganeh, S.; Yost, S. R.; You, Z.-Q.; Zhang, I. Y.; Zhang, X.; Zhao, Y.; Brooks, B. R.; Chan, G. K. L.; Chipman, D. M.; Cramer, C. J.; Goddard III, W. A.; Gordon, M. S.; Hehre, W. J.; Klamt, A.; Schaefer III, H. F.; Schmidt, M. W.; Sherrill, C. D.; Truhlar, D. G.; Warshel, A.; Xu, X.; Aspuru-Guzik, A.; Baer, R.; Bell, A. T.; Besley, N. A.; Chai, J.-D.; Dreuw, A.; Dunietz, B. D.; Furlani, T. R.; Gwaltney, S. R.; Hsu, C.-P.; Jung, Y.; Kong, J.; Lambrecht, D. S.; Liang, W.; Ochsenfeld, C.; Rassolov, V. A.; Slipchenko, L. V.; Subotnik, J. E.; Van Voorhis, T.; Herbert, J. M.; Krylov, A. I.; Gill, P. M. W.; Head-Gordon, M. Advances in molecular quantum chemistry contained in the Q-Chem 4 program package. *Mol. Phys.* **2015**, *113*, 184-215.
- [36] Gaussian 03, Revision C.02, Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, Jr., J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; and Pople, J. A.; Gaussian, Inc., Wallingford CT, 2004.
- [37] Jensen, R. K.; Korcek, S.; Mahoney, L. R.; Zinbo, M. *J. Am. Chem. Soc.* **1979**, *101*, 7574.
- [38] Jalan, A.; Allen, J. W.; Green, W. H. Chemically activated formation of organic acids in reactions of the Criegee intermediate with aldehydes and ketones. *Phys. Chem. Chem. Phys.* **2013**, *15*, 16841-16852.
- [39] Zimmerman, P. M. Reliable Transition State Searches Integrated with the Growing String Method. *J. Chem. Theory Comput.* **2013**, *9*, 3043-3050.
- [40] Zimmerman, P. M. Single-ended transition state finding with the growing string method. *J. Chem. Theory Comput.* **2015**, *36*, 601-611.

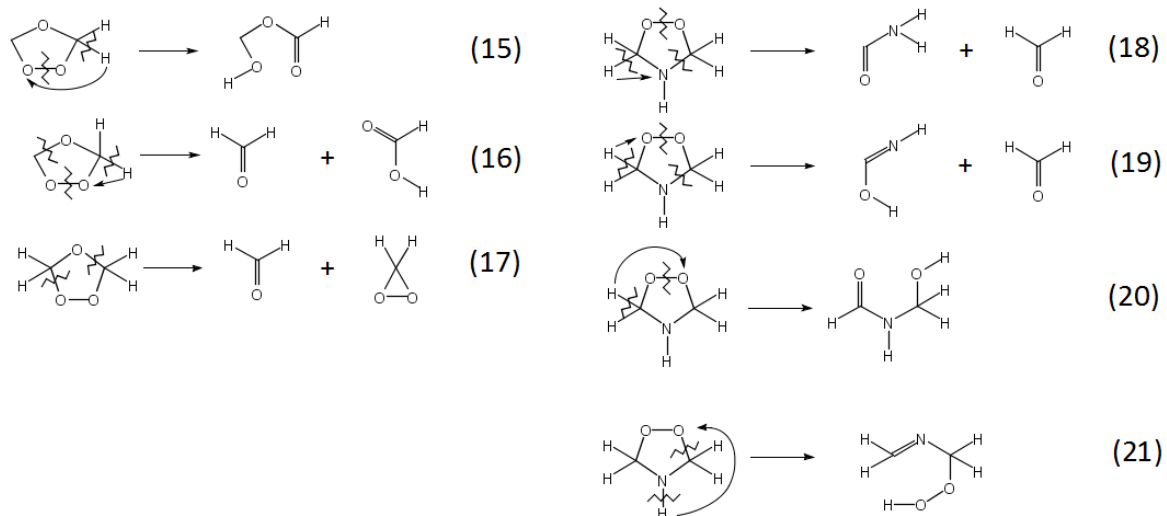
[41] Zimmerman, P. M. Automated Discovery of Chemically Reasonable Elementary Reaction Steps. *J. Comput. Chem.* **2013**, *34*, 1385-1392.

**Figure 1.** Flowchart for automated discovery of elementary chemical reaction steps using molecular mechanics and quantum chemistry calculations. See “Theory” section for details of each step.

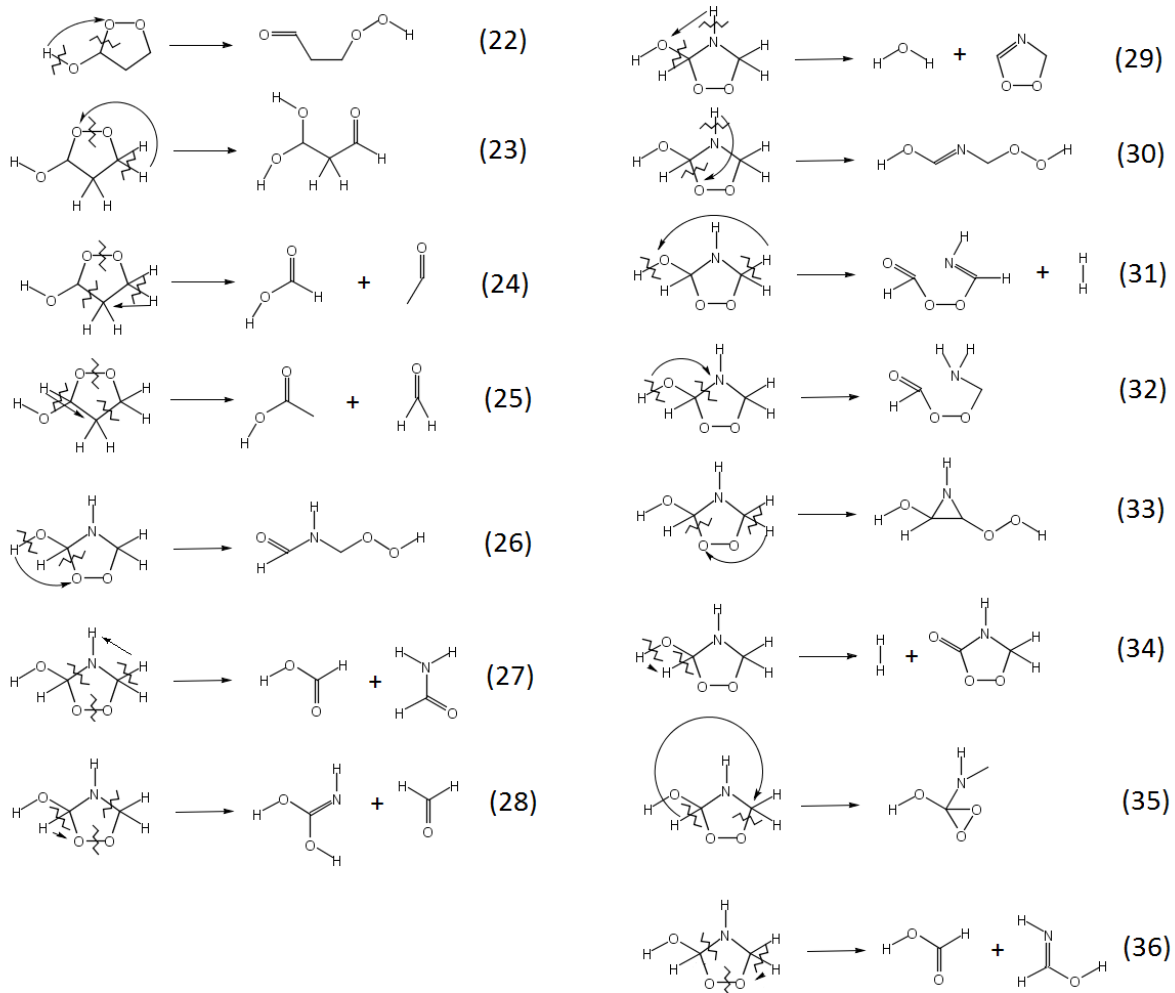




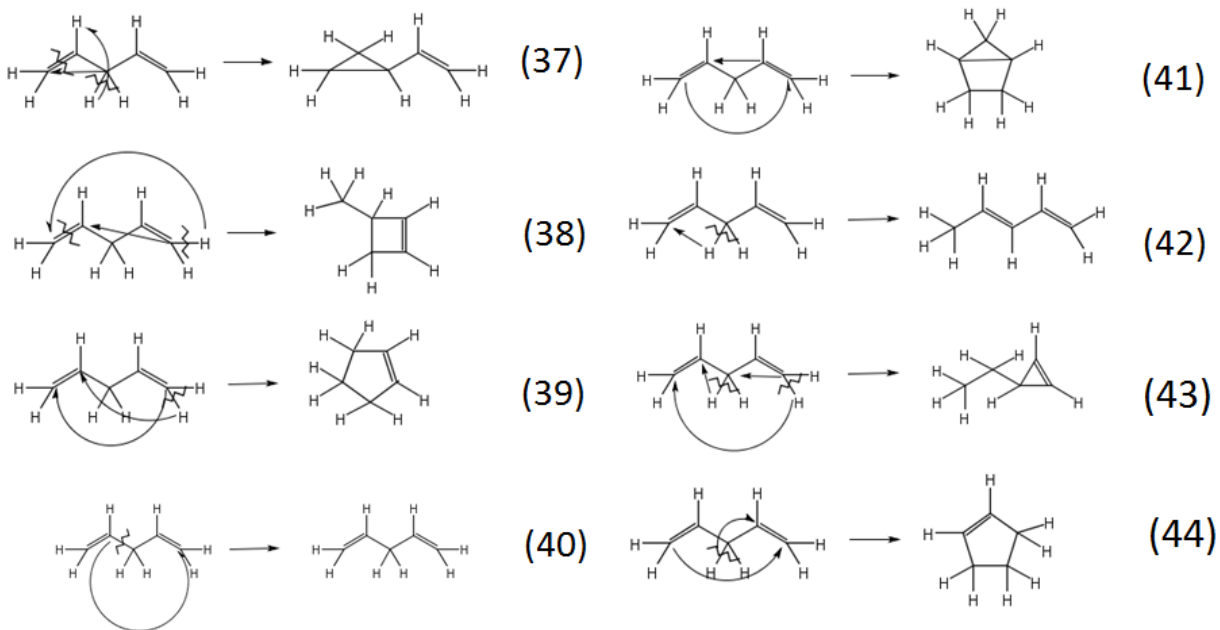
**Figure 2.** Automatically identified unimolecular reactions of  $\gamma$ -ketohydroperoxide ( $\text{OCCCC=O}$ ) (1-6) and N-(hydroperoxymethyl) formamide ( $\text{OOCNC=O}$ ) (7-14). Reactions (3), (4), (9)-(12) are unusual and unexpected.



**Figure 3.** Automatically identified unimolecular reactions of ethylene secondary ozonide (O1COOC1) (15-17) and 1,2,4-dioxazolidine (N1COOC1) (18-21). Reactions (17) and (21) are particularly unusual and unexpected.



**Figure 4.** Automatically identified unimolecular reactions of 1,2-dioxolan-3-ol (OC1OCC1) (22-25) and 1,2,4-dioxazolidin-3-ol (OC1OCCN1) (26-36). Reactions (30)-(33) and (35) are particularly unusual and unexpected.



**Figure 5.** Automatically identified unimolecular reactions of 1,4-pentadiene ( $C=CCC=C$ ) (37-44). Reactions (38)-(40) and (43) are particularly unusual and unexpected.

**Table 1.** Number of elementary steps investigated for different initial reactants.

Initial reactant (SMILES)	Total number of product channels identified	Number of low energy channels		Number of automatically identified transition states (versus existing RMG database) <sup>c</sup>
		( $\Delta H_r^0(298K) \leq 0$ kcal/mol)	( $\Delta H_r^0(298K) \leq 20$ kcal/mol)	
$\gamma$ -ketohydroperoxide (OOCCC=O)	77 <sup>a</sup> 415 <sup>b</sup>	39 <sup>a</sup> 204 <sup>b</sup>	53 <sup>a</sup> 276 <sup>b</sup>	6 (RMG database:1)
N- (hydroperoxymethyl) formamide (OOCNC=O)	62 <sup>a</sup> 293 <sup>b</sup>	22 <sup>a</sup> 84 <sup>b</sup>	32 <sup>a</sup> 123 <sup>b</sup>	8 (RMG database:0)
ethylene ozonide (O1COOC1)	42 <sup>a</sup> 82 <sup>b</sup>	23 <sup>a</sup> 59 <sup>b</sup>	35 <sup>a</sup> 80 <sup>b</sup>	3 (RMG database:0)
1,2,4-dioxazolidine (N1COOC1)	56 <sup>a</sup> 190 <sup>b</sup>	35 <sup>a</sup> 137 <sup>b</sup>	49 <sup>a</sup> 159 <sup>b</sup>	4 (RMG database:0)
1,2-dioxolan-3-ol (OC1OCC1)	90 <sup>a</sup> 533 <sup>b</sup>	40 <sup>a</sup> 233 <sup>b</sup>	64 <sup>a</sup> 344 <sup>b</sup>	4 (RMG database:2)
1,2,4-dioxazolidin-3-ol (OC1OOCN1)	74 <sup>a</sup> 378 <sup>b</sup>	43 <sup>a</sup> 191 <sup>b</sup>	59 <sup>a</sup> 275 <sup>b</sup>	11 (RMG database:0)
penta-1,4-diene (C=CCC=C)	96 <sup>a</sup> 652 <sup>b</sup>	45 <sup>a</sup> 197 <sup>b</sup>	64 <sup>a</sup> 341 <sup>b</sup>	8 (RMG database: 3)

<sup>a</sup> Number of breaking bonds = 2; Number of forming bonds = 2

<sup>b</sup> Number of breaking bonds = 3; Number of forming bonds = 3

<sup>c</sup> from <http://rmg.mit.edu>

1.

**1. Report Type**

Final Report

**Primary Contact E-mail****Contact email if there is a problem with the report.**

whgreen@mit.edu

**Primary Contact Phone Number****Contact phone number if there is a problem with the report**

617-253-4580

**Organization / Institution name**

Massachusetts Institute of Technology

**Grant/Contract Title****The full title of the funded effort.**

Automated discovery of new chemical reactions and accurate calculation of their rates

**Grant/Contract Number****AFOSR assigned control number. It must begin with "FA9550" or "F49620" or "FA2386".**

FA9550-13-1-0065

**Principal Investigator Name****The full name of the principal investigator on the grant or contract.**

Dr. William H. Green

**Program Manager****The AFOSR Program Manager currently assigned to the award**

Dr. Chiping R. Li

**Reporting Period Start Date**

03/01/2013

**Reporting Period End Date**

02/28/2015

**Abstract**

A major challenge in combustion chemistry is to be sure that all of the important reactions have been included in the kinetic model. Experience shows that missing reactions are a major cause of discrepancies between model predictions and experimental performance. Looking at combustion chemistry as motion on a Potential Energy Surface landscape, reactions are the passes that connect valleys representing the different molecules; whether or not two valleys are connected is the key issue determining the topology of the landscape. Conventional kinetic modeling approaches include only "well known" types of reactions in the models; as demonstrated in this project these conventional models omit most of the actual reactions occurring in the engine. Discovering new reactions is very challenging, so historically many of the most important reactions have been discovered by serendipity (either experimentally or by using quantum chemistry). In this project we develop a much more systematic approach, using high performance computing to search for reactions connecting a specified reactant to each possible product, and using quantum chemistry to determine the reaction barriers. The algorithm developed here has 4 key steps: (1) identification of all the potential product channels; (2) Search for a low energy reaction path connecting reactant and each product channel; and (3) refinement of the reaction path to determine the saddle point geometry and barrier height. This new computational approach is found to be the most effective method so far at discovering novel chemical reactions. Using the new algorithm we have discovered 38 new

exothermic or mildly endothermic reactions in this project; about half of these newly-discovered reactions were completely surprising. Possible future research directions to make this promising algorithm even more reliable, computationally efficient, and useful for engine modeling purposes are discussed briefly.

#### **Distribution Statement**

This is block 12 on the SF298 form.

Distribution A - Approved for Public Release

#### **Explanation for Distribution Statement**

If this is not approved for public release, please provide a short explanation. E.g., contains proprietary information.

#### **SF298 Form**

Please attach your SF298 form. A blank SF298 can be found [here](#). Please do not password protect or secure the PDF. The maximum file size for an SF298 is 50MB.

[GreenWH\\_MIT\\_AFOSR-FinalReport\\_28MAY2015\\_SF-298.pdf](#)

**Upload the Report Document. File must be a PDF. Please do not password protect or secure the PDF. The maximum file size for the Report Document is 50MB.**

[GreenWH\\_MIT\\_AFOSR-FinalReport\\_28MAY2015.pdf](#)

**Upload a Report Document, if any. The maximum file size for the Report Document is 50MB.**

**Archival Publications (published) during reporting period:**

**Changes in research objectives (if any):**

**Change in AFOSR Program Manager, if any:**

**Extensions granted or milestones slipped, if any:**

**AFOSR LRIR Number**

**LRIR Title**

**Reporting Period**

**Laboratory Task Manager**

**Program Officer**

**Research Objectives**

**Technical Summary**

**Funding Summary by Cost Category (by FY, \$K)**

	Starting FY	FY+1	FY+2
Salary			
Equipment/Facilities			
Supplies			
Total			

**Report Document**

**Report Document - Text Analysis**

**Report Document - Text Analysis**

**Appendix Documents**

**2. Thank You**

**E-mail user**

May 29, 2015 12:13:30 Success: Email Sent to: whgreen@mit.edu